

# Guidelines for Determining Flood Flow Frequency Bulletin 17C



## Techniques and Methods 4–B5

**U.S. Department of the Interior  
U.S. Geological Survey**

**DRAFT: August 31, 2017**

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

**Cover.** Photograph of international bridge over the St. John River in Fort Kent, Maine at Station 01014000, during flood of April 30, 2008. Photo by M. Huard, USGS.

**DRAFT: August 31, 2017**

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

# **Guidelines for Determining Flood Flow Frequency**

## **Bulletin 17C**

By John F. England, Jr., Timothy A. Cohn, Beth A. Faber, Jery R. Stedinger, Wilbert O. Thomas Jr., Andrea G. Veilleux, Julie E. Kiang, and Robert R. Mason

Techniques and Methods 4–B5

**U.S. Department of the Interior**  
**U.S. Geological Survey**

**U.S. Department of the Interior**  
RYAN ZINKE, Secretary

**U.S. Geological Survey**  
William Werkheiser, Director (Acting)

U.S. Geological Survey, Reston, Virginia: 2017

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment, visit <http://www.usgs.gov> or call 1-888-ASK-USGS

For an overview of USGS information products, including maps, imagery, and publications, visit <http://www.usgs.gov/pubprod>

To order this and other USGS information products, visit <http://store.usgs.gov>

Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this report is in the public domain, permission must be secured from the individual copyright owners to reproduce any copyrighted materials contained within this report.

*Suggested citation:*

England, J.F. Jr., Cohn, T.A., Faber, B.A., Stedinger, J.R., Thomas Jr., W.O., Veilleux, A.G., Kiang, J.E., and Mason, R.R., 2017, Guidelines for Determining Flood Flow Frequency – Bulletin 17C: U.S. Geological Survey Techniques and Methods book 4, chap. B5, XXX p., <https://dx.doi.org/10.3133/tm4-B5/>.

ISSN 2328-7055 (online)



# Contents

Abstract .....	1
Introduction .....	1
Background .....	2
Purpose and Scope .....	3
Risk Accumulates .....	4
Acknowledgments .....	5
Flood Flow Frequency Information .....	5
Use of Annual Maximum Series .....	5
Data Sources for a Site .....	6
Systematic Records .....	6
Historical Flood Information .....	7
Paleoflood and Botanical Information .....	8
Common Issues with At-Site Data Records .....	11
Broken, Incomplete, and Discontinued Records .....	12
Extraordinary Floods .....	12
Zero Flows and <del>Potentially-Influential</del> <u>Potentially Influential</u> Low Floods .....	14
Data Representation using Flow Intervals and Perception Thresholds .....	15
Regional Information and Nearby Sites .....	18
Flood Estimates from Precipitation .....	19
Data Assumptions and Specific Concerns .....	20
Flow Measurement Error .....	20
Randomness of Events .....	21
Mixed Populations .....	22
Watershed Changes .....	23
Climate Variability and Change .....	23
Determination of the Flood Flow Frequency Curve .....	24
Plotting Positions .....	24
Flood Distribution .....	24
Parameter Estimation: Simple Case .....	25
Moments and Parameters .....	25
Weighted Skew Coefficient Estimator .....	26
Zeros and Identifying Potentially Influential Low Floods .....	26
Expected Moments Algorithm .....	27
Record Extension with Nearby Sites .....	29
Confidence Intervals for Quantiles .....	30

Estimating Regional Skew .....	30
Comparisons of Frequency Curves .....	31
Comparisons with Similar Watersheds .....	32
Comparisons with Flood Estimates from Precipitation .....	33
Weighting of Independent Frequency Estimates .....	33
Frequency Curve Extrapolation .....	34
Software and Examples .....	34
Future Studies .....	35
Ungaged Sites .....	35
Regulated Flow Frequency .....	36
Urbanization and Watershed Change .....	36
Applicability of These Guidelines .....	36
References .....	38
Glossary .....	53
Appendix 1. List of Symbols .....	61
List of Symbols .....	63
Appendix 2. Subcommittee and Work Group Members .....	65
Subcommittee and Work Group Members .....	67
Appendix 3. Data Sources and Representation .....	69
Data Sources .....	71
Systematic Records .....	71
Historical Data .....	72
Paleoflood Data and Botanical Information .....	73
Regional Information .....	74
Precipitation and Climate Information .....	74
Data Representation .....	75
Appendix 4. Initial Data Analysis .....	79
Initial Data Analysis .....	81
Visual Inspection: Plot the data .....	81
Autocorrelation .....	81
Trends and shifts .....	81
Statistical tests .....	82
Example: Skokie River near Highland Park, IL .....	82
Appendix 5. Threshold-Exceedance Plotting Positions .....	85
Threshold-Exceedance Plotting Positions .....	87
Appendix 6. Potentially Influential Low Floods .....	89
Potentially Influential Low Floods .....	91
PILF Background and Philosophy .....	91
Computational Details for Identifying PILFs with MGBT .....	92

Appendix 7. Expected Moments Algorithm (EMA) .....	95
Expected Moments Algorithm (EMA) .....	97
EMA Computational Details .....	97
The Generalized Expected Moments Algorithm .....	99
Uncertainty of EMA Moments .....	101
Confidence Intervals with EMA .....	103
Appendix 8. Record Extension with Nearby Sites .....	105
Record Extension with Nearby Sites .....	107
Record Augmentation of the Mean and Variance .....	107
Record Extension with MOVE .....	112
Summary of Procedure .....	114
MOVE Example — Suwanee Creek at Suwanee, Georgia .....	116
Appendix 9. Weighting of Independent Estimates .....	123
Weighting of Independent Estimates .....	125
Weighting Method .....	125
Example .....	126
Appendix 10. Examples .....	127
Flood Frequency Steps and Examples .....	129
Systematic Record Example—Moose River at Victory, Vermont .....	131
EMA Representation of Peak Flow Data for Flood Frequency Analysis .....	131
Results from Flood Frequency Analysis .....	131
PILF Example—Orestimba Creek near Newman, California .....	137
EMA Representation of Peak Flow Data for Flood Frequency Analysis .....	137
Results from Flood Frequency Analysis .....	139
Broken Record Example—Back Creek near Jones Springs, West Virginia .....	143
EMA Representation of Peak Flow Data for Flood Frequency Analysis .....	144
Results from Flood Frequency Analysis .....	144
Historical Record Example—Arkansas River at Pueblo, Colorado .....	149
EMA Representation of Peak Flow Data for Flood Frequency Analysis .....	149
Results from Flood Frequency Analysis .....	153
Crest Stage Gage Example—Bear Creek at Ottumwa, Iowa .....	155
EMA Representation of Peak Flow Data for Flood Frequency Analysis .....	155
Results from Flood Frequency Analysis .....	158
Historical and PILF Example—Santa Cruz River near Lochiel, Arizona .....	161
EMA Representation of Peak Flow Data for Flood Frequency Analysis .....	161
Results from Flood Frequency Analysis .....	163
Paleoflood Record Example—American River at Fair Oaks, California .....	167
EMA Representation of Peak Flow Data for Flood Frequency Analysis .....	167
Results from Flood Frequency Analysis .....	172

## Figures

1. Photograph of a streamflow-gaging station with crest-stage gage at U.S. Geological Survey station 01589238, Gwynns Falls Tributary at McDonogh, Maryland .....	7
2. Photograph of historic flood high-water marks and flood of March 13–15, 2010, Potomac River at Great Falls Park, Virginia .....	7
3. Graph showing an example peak-discharge time series with historical period and discharge perception threshold .....	9
4. Graph showing an example site with three large historical floods outside the gaging record, Big Sandy River at Bruceton, Tennessee .....	10
5. Graph showing an example site with an extraordinary flood peak that represents a longer time frame, Plum Creek near Louviers, Colorado .....	10
6. Graph showing an example site with historical/paleoflood non-exceedance information, North Platte River near Seminoe Reservoir, Wyoming .....	10
7. Diagram of a section showing typical paleoflood features used as paleostage indicators (from Jarrett and England, 2002) .....	11
8. Photograph of inclined western juniper trees with upright branches from 1861–1862 flood, Crooked River near Prineville, Oregon .....	12
9. Graph showing an example streamflow-gaging station with a broken record, U.S. Geological Survey station 01614000, Back Creek near Jones Springs, West Virginia .....	13
10. Graph showing an example streamflow-gaging station with a discontinued record, U.S. Geological Survey station 07099500, Arkansas River near Pueblo, Colorado .....	14
11. Graph showing an example empirical frequency distribution and potentially influential low floods (PILFs) at U.S. Geological Survey station 11274500, Orestimba Creek near Newman, California .....	15
12. Graph showing an example peak-discharge time series with peak flows, interval and binomial-censored flood observations, flow intervals, and perception thresholds .....	17
3.1. Graph showing typical flow intervals and perception thresholds for gaging stations .....	77
3.2. Graph showing typical flow intervals and perception thresholds for crest-stage gages .....	77
3.3. Graph showing typical flow intervals and perception thresholds for historical data .....	78
4.1. Graph showing U.S. Geological Survey station 05535070 Skokie River near Highland Park, Illinois time-series plot .....	83
4.2. Graph showing U.S. Geological Survey station 05535070 Skokie River near Highland Park, Illinois autocorrelation plot with thresholds for significant autocorrelation .....	84

8.1. Map showing location of Suwanee Creek at Suwanee, Georgia, U.S. Geological Survey station 02334885 .....	116
8.2. Graph showing annual peak flows for Suwanee Creek, U.S. Geological Survey station 02334885, the short record station .....	117
8.3. Graph showing annual peak flows for the Etowah River, U.S. Geological Survey station 02392000, the long record station .....	119
8.4. Graph of 20 concurrent years of record for Suwanee Creek and the Etowah River for the period 1985 to 2004 with the ordinary least squares regression line .....	120
10.1. Graph showing U.S. Geological Survey gage 01134500 Moose River at Victory, Vermont annual peak-flow time series consisting of 68 peaks from 1947 to 2014 .....	134
10.2. Graph showing annual exceedance probability plot for U.S. Geological Survey gage 01134500 Moose River at Victory, Vermont based on flood frequency analysis using Expected Moments Algorithm with Multiple Grubbs-Beck Test and weighted skew .....	135
10.3. Graph showing annual exceedance probability plot for U.S. Geological Survey gage 01134500 Moose River at Victory, Vermont based on flood frequency analysis using Expected Moments Algorithm with Multiple Grubbs-Beck Test and station skew only .....	136
10.4. Graph showing U.S. Geological Survey gage 11274500 Orestimba Creek near Newman, California annual peak-flow time series consisting of 82 peaks from 1932 to 2013 .....	137
10.5. Graph showing annual exceedance probability plot for USGS Gage 11274500 Orestimba Creek based on flood frequency analysis using Expected Moments Algorithm with Multiple Grubbs-Beck Test .....	141
10.6. Graph showing U.S. Geological Survey gage 01614000 Back Creek near Jones Springs, West Virginia annual peak-flow time series consisting of 56 peaks from 1929 to 2012 .....	143
10.7. Graph showing annual exceedance probability plot for USGS Gage 01614000 Back Creek near Jones Springs, West Virginia based on flood frequency analysis using Expected Moments Algorithm with Multiple Grubbs-Beck Test .....	145
10.8. Graph showing peak discharge, historical and paleoflood estimates, Arkansas River at Pueblo State Park .....	151
10.9. Graph showing the peak discharge frequency curve, Arkansas River at Pueblo State Park, including gage, historical, and paleoflood data .....	154
10.10. Graph showing U.S. Geological Survey gage 05489490 Bear Creek at Ottumwa, Iowa annual peak-flow time series consisting of 49 peaks from 1965 to 2014 .....	158
10.11. Graph showing annual exceedance probability plot for USGS Gage 05489490 Bear Creek at Ottumwa, Iowa based on flood frequency analysis using Expected Moments Algorithm with Multiple Grubbs-Beck Test .....	159

10.12. Graph showing U.S. Geological Survey gage 09480000 Santa Cruz River near Lochiel, Arizona annual peak-flow time series consisting of 65 peaks from 1949 to 2013 .....	161
10.13. Graph showing annual exceedance probability plot for USGS Gage 09480000 Santa Cruz River based on flood frequency analysis using Expected Moments Algorithm with Multiple Grubbs-Beck Test .....	165
10.14. Graph showing approximate unregulated peak discharge and paleoflood estimates, with historical and paleoflood exceedance thresholds, American River at Fair Oaks .....	169
10.15. Graph showing approximate unregulated peak discharge frequency curve, American River at Fair Oaks, including gage, historical and paleoflood data .....	172

## Tables

1. Flow and year terms used in the Expected Moments Algorithm .....	29
1.1. List of symbols used in this report .....	64
2.1. Subcommittee on Hydrology members .....	67
2.2. Hydrologic Frequency Analysis Work Group members .....	68
3.1. Generalized data representation of peak-flow intervals and perception thresholds for each year .....	76
5.1. Typical plotting position parameter $a$ values .....	88
7.1. Expected Moments Algorithm censored-data threshold categories .....	102
8.1. Summary of concurrent observed annual peak data for the Etowah River and Suwanee Creek .....	118
8.2. MOVE extended record for 13 years (1972–1984) for Suwanee Creek at Suwanee, Georgia .....	121
8.3. Flood records for 93 years for the Etowah River at Canton, Georgia .....	122
10.1. Summary of flood frequency examples .....	130
10.2. U.S. Geological Survey gage 01134500 Moose River at Victory, Vermont annual peak-flow record consisting of 68 peaks from 1947 to 2014 .....	132
10.3. U.S. Geological Survey gage 01134500 Moose River at Victory, Vermont Expected Moments Algorithm flow intervals for the systematic period from 1947 to 2014 .....	133
10.4. U.S. Geological Survey gage 01134500 Moose River at Victory, Vermont Expected Moments Algorithm perception thresholds for the systematic period from 1947 to 2014 .....	134

10.5. Peak-flow quantiles in cubic feet per second for U.S. Geological Survey-gage 01134500 Moose River at Victory, Vermont based on flood frequency analysis using Expected Moments Algorithm with Multiple Grubbs-Beck Test .....	135
10.6. U.S. Geological Survey gage 11274500 Orestimba Creek near Newman, California annual peak-flow record from 1932 to 2013 .....	138
10.7. U.S. Geological Survey gage 11274500 Orestimba Creek Expected Moments Algorithm flow intervals from 1932 to 2013 .....	140
10.8. U.S. Geological Survey gage 11274500 Orestimba Creek near Newman, California Expected Moments Algorithm perception thresholds for the period from 1932 to 2013 .....	141
10.9. Peak-flow quantiles for USGS Gage 11274500 Orestimba Creek near Newman, California based on flood frequency analysis using Expected Moments Algorithm with Multiple Grubbs-Beck Test .....	142
10.10. U.S. Geological Survey gage 01614000 Back Creek near Jones Springs, West Virginia annual peak-flow record consisting of 56 peaks from 1929 to 2012 .....	146
10.11. U.S. Geological Survey gage 01614000 Back Creek near Jones Springs, West Virginia Expected Moments Algorithm flow intervals for the systematic period from 1929 to 2012 .....	147
10.12. U.S. Geological Survey gage 01614000 Back Creek near Jones Springs, West Virginia Expected Moments Algorithm perception thresholds for the systematic period from 1929 to 2012 .....	148
10.13. Peak-flow quantiles in cubic feet per second for USGS Gage 01614000 Back Creek near Jones Springs, West Virginia based on flood frequency analysis using Expected Moments Algorithm with Multiple Grubbs-Beck Test .....	148
10.14. U.S. Geological Survey gage 07099500 (and others) Arkansas River annual peak-flow record consisting of 85 peaks from 1864 to 1976 .....	150
10.15. Arkansas River at Pueblo EMA flow intervals for the period from 1864 to 1976 .....	152
10.16. USGS gage 07099500 (and others) Arkansas River Expected Moments Algorithm perception thresholds for the historical, paleoflood and systematic period from 1165 to 2004 .....	153
10.17. Peak-flow quantiles for U.S. Geological Survey gage 07099500 (and others) Arkansas River based on flood frequency analysis using Expected Moments Algorithm with Multiple Grubbs-Beck Test .....	154
10.18. U.S. Geological Survey gage 05489490 Bear Creek at Ottumwa, Iowa annual peak-flow record consisting of 49 peaks from 1965 to 2014 .....	156
10.19. U.S. Geological Survey gage 05489490 Bear Creek at Ottumwa, Iowa Expected Moments Algorithm flow intervals for the systematic period from 1965 to 2014 .....	157
10.20. U.S. Geological Survey gage 05489490 Bear Creek at Ottumwa, Iowa Expected Moments Algorithm perception thresholds for the systematic period from 1965 to 2014 .....	159

10.21.	Peak-flow quantiles in cubic feet per second for USGS Gage 05489490 Bear Creek at Ottumwa, Iowa based on flood frequency analysis using Expected Moments Algorithm with Multiple Grubbs-Beck Test .....	160
10.22.	U.S. Geological Survey gage 09480000 Santa Cruz River near Lochiel, Arizona peak-flow record consisting of 65 peaks from 1949 to 2013 .....	162
10.23.	U.S. Geological Survey gage 09480000 Santa Cruz River near Lochiel, Arizona Expected Moments Algorithm flow intervals for the systematic period from 1949 to 2013 .....	164
10.24.	U.S. Geological Survey gage 09480000 Santa Cruz River near Lochiel, Arizona Expected Moments Algorithm perception thresholds for the his- torical and systematic period from 1927 to 2013 .....	165
10.25.	Peak-flow quantiles for USGS Gage 09480000 Santa Cruz River near Lochiel, Arizona based on flood frequency analysis using Expected Moments Algorithm with Multiple Grubbs-Beck Test .....	166
10.26.	U.S. Geological Survey gage 11446500 American River at Fair Oaks annual peak-flow record with historical floods and paleofloods .....	168
10.27.	U.S. Geological Survey gage 11446500 American River at Fair Oaks Expected Moments Algorithm flow intervals for the historical and systematic period from 650 to 1997 .....	170
10.28.	U.S. Geological Survey gage 11446500 American River at Fair Oaks Expected Moments Algorithm perception thresholds for the paleoflood, histori- cal, and systematic period from 1 to 2000 .....	171
10.29.	Peak-flow quantiles in cubic feet per second for USGS Gage 11446500 American River at Fair Oaks based on flood frequency analysis using Expected Moments Algorithm with Multiple Grubbs-Beck Test .....	173



## Conversion Factors

Multiply	By	To obtain
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
square mile (mi <sup>2</sup> )	2.590	square kilometer (km <sup>2</sup> )
cubic foot per second (ft <sup>3</sup> /s)	0.02832	cubic meter per second (m <sup>3</sup> /s)

## Datum

Horizontal coordinate information is referenced to the North American Datum of 1983 (NAD 83).  
Vertical coordinate information is referenced to the National Geodetic Vertical Datum of 1929 (NGVD 29).

## Abbreviations

**ACWI**—Advisory Committee on Water Information  
**AEP**—annual exceedance probability  
**AMS**—annual maximum series  
**B-GLS**—Bayesian generalized least squares  
**CDF**—cumulative distribution function  
**CSG**—Crest-stage gage  
**EMA**—Expected Moments Algorithm  
**ENSO**—El Niño-Southern Oscillation  
**FEMA**—Federal Emergency Management Agency  
**GLS**—generalized least squares  
**HFAWG**—Hydrologic Frequency Analysis Work Group  
**HWM**—High-water mark  
**LP-III**—Log-Pearson Type III distribution  
**MGBT**—Multiple Grubbs-Beck Test  
**MOVE**—Maintenance of Variance Extension  
**MSE**—mean square error  
**NRCS**—Natural Resources Conservation Service  
**NWIS**—USGS National Water Information System  
**NWS**—National Weather Service  
**OLS**—ordinary least squares  
**PDS**—partial-duration series  
**PILF**—Potentially Influential Low Flood  
**PSI**—paleostage indicator  
**Reclamation**—Bureau of Reclamation  
**RFC**—River Forecast Center (NWS)  
**SOH**—Subcommittee on Hydrology  
**USACE**—U.S. Army Corps of Engineers  
**USGS**—U.S. Geological Survey  
**WLS**—weighted least squares



# Guidelines for Determining Flood Flow Frequency

## Bulletin 17C

By John F. England, Jr., Timothy A. Cohn, Beth A. Faber, Jerry R. Stedinger, Wilbert O. Thomas Jr., Andrea G. Veilleux, Julie E. Kiang, and Robert R. Mason

### Abstract

Accurate estimates of flood frequency and magnitude are a key component of any effective nationwide flood risk management and flood damage abatement program. In addition to accuracy, methods for estimating flood risk must be uniformly and consistently applied because management of the Nation's water and related land resources is a collaborative effort involving multiple actors including most levels of government and the private sector.

Flood frequency guidelines have been published in the United States since 1967, and have undergone periodic revisions. In 1967, the U.S. Water Resources Council presented a coherent approach to flood frequency with Bulletin 15 (USWRC, 1967), "A Uniform Technique for Determining Flood Flow Frequencies." The method it recommended involved fitting the log-Pearson Type III distribution to annual peak flow data by the ~~method-of-moments~~method of moments. The first extension and update of Bulletin 15 was published in 1976 as Bulletin 17 (USWRC, 1976), "Guidelines for Determining Flood Flow Frequency" (Guidelines). It extended the Bulletin 15 procedures by introducing methods for dealing with outliers, historical flood information, and regional skew. Bulletin 17A was published the following year to clarify the computation of weighted skew. The next revision of the Bulletin, 17B (IACWD, 1982), provided a host of improvements and new techniques designed to address situations that often arise in practice, including better methods for estimating and using regional skew, weighting station and regional skew, detection of outliers, and use of the conditional probability adjustment (CPA)

(Thomas, 1985; Griffis and Stedinger, 2007a).

The current version of the Guidelines are presented in this document, denoted Bulletin 17C. It incorporates changes motivated by four of the items listed as "Future Work" in Bulletin 17B and 30 years of post-17B research on flood processes and statistical methods. The updates include: adoption of a generalized representation of flood data that allows for interval and censored data types; a new method, called the Expected Moments Algorithm (Cohn and others, 1997, 2001), that extends the ~~method-of-moments~~method of moments so that it can accommodate interval data; a generalized approach to identification of low outliers in flood data (Cohn and others, 2013); and an improved method for computing confidence intervals.

Federal agencies are requested to use these guidelines in all planning activities involving water and related land resources. State, local, and private organizations are encouraged to use these guidelines to assure uniformity in the flood frequency estimates that all agencies concerned with flood risk should use for Federal planning decisions.

This revision is adopted with the knowledge and understanding that review of these procedures will be ongoing. Updated methods will be adopted when warranted by experience and by examination and testing of new techniques.

### Introduction

These Guidelines describe the data and procedures for computing flood flow frequency where sys-

PROVISIONAL  
THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.  
IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).  
IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

## 2 Guidelines for Determining Flood Flow Frequency – Bulletin 17C

tematic ~~stream-gaging~~ streamgaging records of sufficient length (at least 10 years, with an informative regional skew and ~~for~~(or) record extension) to warrant statistical analysis are available. The procedures do not cover watersheds where flood flows are appreciably altered by reservoir regulation, watershed changes, or hydrologic nonstationarities, or where the possibility of unusual events, such as dam failures, must be considered.

### Background

In December 1967, Bulletin No. 15, “A Uniform Technique for Determining Flood Flow Frequencies”,<sup>1</sup> was issued by the Hydrology Committee of the Water Resources Council (USWRC, 1967). The report recommended use of the Pearson Type III distribution with log transformation of the data (log-Pearson Type III distribution) as a base method for flood flow frequency studies. As pointed out in that report, further studies were needed covering various aspects of flow frequency determinations.

In March 1976, Bulletin 17, “Guidelines for Determining Flood Flow Frequency” was issued by the Water Resources Council (USWRC, 1976). The guide was an extension and update of Bulletin No. 15. It provided a more complete guide for flood flow frequency analysis, incorporating currently accepted technical methods with sufficient detail to promote uniform application. It was limited to defining flood potentials in terms of peak discharge and exceedance probability at locations where a systematic record of peak flood flows is available. The recommended set of procedures was selected from those used or described in the literature prior to 1976, based on studies conducted for this purpose at the Center for Research in Water Resources of the University of Texas at Austin (Beard, 1974) that are summarized in IACWD (1982, ~~Appendix 14~~) IACWD (1982, appendix 14) and in other studies by the Work Group on Flood Flow Frequency.

The Guidelines were revised and reissued in June 1977 as Bulletin 17A, which clarified the procedure for computing weighted skew. Bulletin 17B was the next effort to improve and expand upon the earlier publications. Bulletin 17B was issued in 1981, and ~~re-issued~~

reissued with minor corrections in 1982 (IACWD, 1982). Bulletin 17B provided revised procedures for weighting station skew values with results from a generalized skew study, detecting and treating outliers, making two station comparisons, and computing confidence limits about a frequency curve. Thomas (1985) and Griffis and Stedinger (2007a) present additional details on the history of Bulletins 17, 17A, and 17B.

In 2005, the Hydrologic Frequency Analysis Work Group (HFAWG), under the Subcommittee on Hydrology (SOH), began discussing recent research on flood frequency and potential significant revisions to Bulletin 17B. The HFAWG submitted a plan to SOH in 2006 (Hydrologic Frequency Analysis Work Group, 2006) to conduct studies on flood frequency improvements. The focus was on evaluating a generalized method of moments approach (Cohn and others, 1997), with tests on gaging station peak-flow data and with Monte-Carlo simulation (Cohn and others, 2017). New procedures were developed to deal with troublesome ~~data-sets~~ datasets, and new methods were extensively tested with selected ~~data-sets~~ datasets and in Monte Carlo studies (Cohn and others, 2017). In 2013, the HFAWG made recommendations to SOH to revise Bulletin 17B (Hydrologic Frequency Analysis Work Group, 2013). Additional background on revision efforts is available on the HFAWG ~~webpage~~ web page at <https://acwi.gov/hydrology/Frequency/minutes/index.html>. ~~Appendix 2 lists HFAWG and SOH members involved in the study and revision effort.~~

This document is an update to the guidelines published earlier in Bulletins 17, 17A, and 17B. Revisions incorporated in this document address major limitations of Bulletin 17B. Most of these limitations were well known, and are listed in Bulletin 17B (IACWD, 1982) on ~~pp. 27–28~~ p. 27–28 as topics needing future study.

A particularly important innovation in these new guidelines is the elimination of the need, implicit in application of Bulletin 17B, that all annual peaks be either point-value flow estimates, or upper bounds on historical flows, or on ~~low-flows~~ low flows and zeros. With new statistical and computational procedures, these Guidelines employ a new comprehensive data framework; flood data are now generalized as “interval estimates” that incorporate both standard

PROVISIONAL  
THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

point-value flood observations, as well as upper bound, lower bounds, or simple interval estimates describing the value of the peak flood in each year.

These Guidelines take advantage of the new data framework by utilizing the Expected Moments Algorithm (EMA) to analyze available flood data in a single, uniform, and consistent framework that does not require the introduction of additional algorithms to adjust the flood frequency curve to incorporate or account for the presence in the dataset of historic information, zero flows, or low outliers, as is the case with Bulletin 17B. Thus, it avoids the need for arbitrary selection of a sequence of such adjustments described on pages ~~12-2 through 12-4~~ 12-2 through 12-4 of Bulletin 17B.

These Guidelines improve on Bulletin 17B by introducing a standardized Multiple Grubbs-Beck test to identify potentially influential low flood observations (PILFs), which can be given special treatment to prevent their exerting excessive influence on the fitting of the flood frequency curve. This is a very important addition because the new procedure provides clear, ~~reasonable and an~~, and objective steps for the identification of such PILFs.

In addition, these Guidelines improve on procedures for estimating regional skewness estimators and their precision, thus replacing the map provided in ~~Plate plate~~ 1 of Bulletin 17B. The recommended procedure employs Bayesian Generalized Least Squares (GLS) regionalization concepts to develop improved estimates of regional flood skew, reflecting the precision of available estimates, their ~~cross-correlation~~ cross correlation, and the precision of the regional model.

Finally, taking together the use of the interval-data framework, EMA, and Bayesian skew coefficient regionalization permits the development of a more accurate estimation of confidence intervals about the flood frequency curve than do procedures described in Bulletin 17B. Large differences in confidence intervals may be observed between intervals computed with Bulletin 17B and procedures in these Guidelines (Bulletin 17C) because the Bulletin 17B confidence intervals ignored uncertainty in the estimated skewness coefficient, ~~and~~ had no provision for recognizing the value of historical information.

## Purpose and Scope

The present Guidelines incorporate updated flood frequency methods based on research summarized by Stedinger and Griffis (2008), concepts described by England and Cohn (2007, 2008), testing by Cohn and others (2017), and a substantial body of literature over the past 30 years cited throughout this document (see References). These updated methods address some of the recommended research and limitations in Bulletin 17B. The ~~following~~ important improvements include the following:

1. the ability to accommodate a generalized form of peak-flow data, specifically interval estimates of ~~peak-discharge~~ peak-discharge magnitudes;
2. a generalization of the ~~method-of-moments~~ method of moments that can accommodate interval, censored, and binomial-censored data called the Expected Moments Algorithm (EMA) (Cohn and others, 1997);
3. accurate confidence interval formulas that can account for historical and paleoflood information as well as regional skew information (Cohn and others, 2001; Cohn, 2015); and
4. a generalized low-outlier procedure, based on the existing Grubbs-Beck test, called the Multiple Grubbs-Beck Test (MGBT), ~~that which~~ that can identify multiple potentially ~~influential~~ influential low floods in the peak flow dataset (Cohn and others, 2013).

These Guidelines are divided into nine sections, which are summarized below.

**Flood Flow Frequency Information –** ~~The following categories of flood~~ Flood data are recognized in the following categories: systematic records, historical data, paleoflood and botanical data, regional information, comparison with similar watersheds, and flood estimates from precipitation. Common data issues and representation of data using intervals and thresholds are presented. How each can be used to define the flood potential is briefly described.

**Data Assumptions and Specific Concerns –** A brief discussion of basic data assumptions is presented

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

#### 4 Guidelines for Determining Flood Flow Frequency – Bulletin 17C

as a reminder to those developing flood flow frequency curves to be aware of potential data issues and concerns. Flow measurement error, randomness of events, trends, long-term persistence, mixed populations, watershed changes, and climate variability are briefly discussed.

##### Determination of the Flood Flow Frequency

##### Curve – This section provides guidance for the determination of a frequency curve. The Pearson Type

III distribution with log transformation of the flood

data (log-Pearson Type III) is recommended as the

basic distribution for defining the annual flood series

(USWRC, 1967; IACWD, 1982; Griffis and Stedinger,

2007b). The method of moments with the ~~Expected~~

~~Moments Algorithm~~-EMA is used to estimate the

parameters of the distribution from station data, includ-

ing historical and paleoflood data, when available.

Adjustments are made for potentially ~~-influential~~ influential

low floods. Regional information is used to estimate

the skew coefficient. Optional record extension meth-

ods using nearby stations is presented. Statistical

uncertainty in flood quantile estimates, including the

construction of confidence interval, is described.

**Estimating Regional Skew –** The general procedure

that is recommended to estimate a regional skew

is described.

**Comparisons of Frequency Curves –** Some concepts

are described for making comparisons of frequency

curves estimated using the procedures in this

guide to those from similar watersheds and flood

estimates from precipitation. In some situations, a

weighted combination of frequency curves may provide

an improved estimate.

**Software and Examples –** Software to estimate

frequency curves and examples demonstrating the use

of these procedures are described.

**Future Studies –** Recommended future studies

are listed, including methods for ungaged sites and for

regulated frequency and urbanization situations.

**Applicability of These Guidelines –** The applicability

of these Guidelines and some limitations are

discussed in this section.

**Appendix –** The appendixes provide information

on data sources, procedures for initial data analysis,

the methods and some computational details for

the recommended procedures, flood frequency steps

and examples that implement the recommended procedures,

~~and a.~~ Symbols are defined in appendix 1.

Appendix 2 lists HFAWG and SOH members involved

in the study and revision effort.

It is possible to standardize many elements of

flood frequency analysis. These Guidelines describe

each major element of the process of defining the

flood potential at a specific location in terms of peak

discharge and annual exceedance probability (AEP).

Flood quantiles with AEP ranging from 0.10 to about

0.002 are estimated using annual maximum flood

series and methods described here. These estimates

depend on the data used in the analysis. When longer

historical and paleoflood records are used ( $>1,000$

years), floods with ~~AEPs  $<0.002$~~  AEPs  $<0.002$  can

be estimated. Use is confined to stations where available

records are adequate to allow reliable statistical

analysis of the data. Special situations may require

other approaches. In those cases where the procedures

of this guide are not followed, deviations must

be supported by appropriate study and accompanied

by a comparison of results using the recommended

procedures.

Flood records are limited. As more years of record

become available at each location, the determination

of flood potential may change. Thus, an estimate

may be outdated a few years after it is made. Additional

flood data alone may be sufficient reason for a

fresh assessment of the flood potential. When making

a new assessment, the analyst should incorporate ~~in~~

~~their study~~ a review of earlier estimates in their study.

Where differences appear, they should be acknowledged

and explained.

**Risk Accumulates**

It is important to realize that the probabilities computed

here correspond to the annual exceedance probability,

or the probability in any year that a flood threshold

is exceeded. However, when considering the chance

that homes, stores, factories, and other public

and private facilities are flooded, owners and occupants

should consider the likelihood of flooding not

just in a single year, but the chance over 10, 25, or

even 100 years. Such permanent facilities are generally

built with design lives ~~(corresponding to a plan-~~

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017



ning horizon, ~~7~~) of 25 or more years.

As used in this guide, risk is defined as the probability that one or more events will exceed a given flood magnitude within a specified period of years  $n$ . Assuming the flow frequency curve is accurate and that events from ~~year-to-year~~ year to year are independent, the probability  $p_n$  that a damage threshold is exceeded at least once in an  $n$ -year period is (Yen, 1970; Kite, 1988), as shown in the following equation:

$$p_n = 1 - (1 - p)^n \quad (1)$$

where  $p$  is the annual exceedance probability (AEP) for each year.

Thus, given the probability that a threshold has an AEP of 0.01 (or 1%), over a ~~25-year period~~ 25-year period, there is a 22% chance of the threshold being exceeded, ~~over a 50-year period;~~ over a 50-year period, there is a 40% chance of the threshold being exceeded, and over a ~~100-year period~~ 100-year period, there is a 63% chance of the threshold being exceeded. Or viewed another way, a new home or business that is protected to have only a 1% chance of being flooded in a single year has a 26% probability of being flooded over the life of a 30-year mortgage. Thus, there is a ~~1-in-4~~ one-in-four chance the property will be flooded in that time period. While the probability of flooding in a single year may seem small when the AEP is just 1% or less, the chance of flooding accumulates over time so that the probability of flooding over 25 or 50 years is substantial. A full risk analysis that includes uncertainty (National Research Council, 2000) is an addition that could be considered, but is beyond the scope of these Guidelines.

## Acknowledgments

These revised Guidelines were written by a team assembled from the Hydrologic Frequency Workgroup (HFAWG), under the auspices of the Subcommittee on Hydrology (SOH) of the Advisory Committee on Water Information (ACWI). Input from other members of HFAWG and of the SOH is gratefully acknowledged. HFAWG and SOH Work Group Members and participants in this revision are listed in appendix 2. The authors

would also like to acknowledge the contributions of the USGS colleague reviewers, whose comments and suggestions materially improved this report. We appreciate the assistance of Dr. William H. Asquith (USGS) with LaTeX and typesetting of this report.

Lastly, our colleague and friend, Dr. Timothy A. Cohn, authored or coauthored many of the papers the form the foundation of this update to the flood frequency guidelines. He wrote software that was critical to implementing the concepts and shared it with others. He gave freely of his time in teaching and instructing others in flood frequency and statistics. Tim was also instrumental in building support amongst HFAWG, SOH, and many Federal agencies. He will be missed.

Timothy A. Cohn

February 26, 1957 — February 20, 2017

## Flood Flow Frequency Information

When developing a flood flow frequency curve, the analyst should consider all available information. The general types of data and information ~~which that~~ can be included in the flood flow frequency analysis are described in the following sections, as well as how to best characterize available data. Flood frequency analysis relies primarily on systematic records, which typically can be represented as point observations. Other types of data, such as historical and paleoflood data, may be represented with intervals or thresholds, because the magnitudes of flood peaks might be known with less precision. The analyst also needs to consider the use of regional information and flood estimates from precipitation. Specific applications are discussed in subsequent sections of this guide.

## Use of Annual Maximum Series

Flood events can be analyzed using either annual maximum series (AMS) or partial-duration series (PDS). The annual maximum flood series is based on the instantaneous maximum flood peak for each year. Annual maximum mean daily discharge or annual maximum ~~n-day~~ provisional flood volumes (U.S. Army Corps of

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

Engineers, 1993; Lamontagne and others, 2012) may also be considered, depending on the intended use of the flood frequency relationship. A ~~partial-duration series~~PDS is obtained by taking all flood peaks equal to or greater than a predefined base flood. Thus an  $n$ -year record can produce  $m$  peaks with  $m > n$ .

Flood frequency estimates using these Guidelines are appropriate for the 0.10 AEP or less flood ( $Q_p > Q_{0.10}$ ), where the AEP  $p \leq 0.1$ , such as 0.01. The annual maximum flood series provides a satisfactory sample for this type of analysis. There is little difference in AEP estimates using AMS or PDS for these quantiles (Langbein, 1949). The AMS is also used due to widespread availability and extended length of AMS data. There are limited PDS records and challenges in defining PDS threshold(s) (Madsen and others, 1997).

If minor floods are of interest, with  $Q_p \leq Q_{0.10}$  AEP, a ~~partial-duration series~~PDS may be appropriate. The PDS base is selected to assure that all events of interest are evaluated. A major problem encountered when using a ~~partial-duration series~~PDS is to define flood events to ensure that all events are independent. It is common practice to establish an empirical basis for separating flood events (Lang and others, 1999). The basis for separation will depend upon the investigator and the intended use. No specific guidelines are recommended for defining flood events to be included in a partial series.

Beard (1974) sought to determine if a consistent relationship existed between the annual and partial series, which could be used to convert from the annual to the ~~partial-duration series~~PDS. Based on that work, it is recommended that the partial-duration series be developed from observed data. An alternative but less desirable solution is to convert from the annual to the ~~partial-duration series~~PDS using a factor.

The procedures described in this guide apply to the annual maximum flood series. If minor flood estimates are needed ( $Q_p \leq Q_{0.10}$ ), such as  $Q_{0.95}$ , a frequency analysis such as ~~peaks-over-threshold~~peaks over threshold (Stedinger and others, 1993; Coles, 2001) using partial-duration data may be appropriate. No specific guidelines are recommended for conducting a partial-duration frequency analysis.

## Data Sources for a Site

The main data sources that are recommended for use in flood frequency include systematic records, historical flood information, and paleoflood and botanical information. These at-site flood data are briefly described; additional information on data sources is in ~~Appendix~~appendix 3. Refer to the Glossary for data-related definitions and notation.

## Systematic Records

Systematic flood data consist of annual peak discharge data collected at regular, prescribed intervals at a gaging station (Salas and others, 1994; Wahl and others, 1995). Systematic flood data involve the continuous monitoring of flood properties by hydrologists (Rantz and Others, 1982a; Baker, 1987). In the United States, the U.S. Geological Survey operates and maintains a nationwide gaging station network (Wahl and others, 1995), and is the primary source for systematic flood data. Stream gages are also operated by ~~federal agencies~~(e.g. Federal agencies (for example, Bureau of Reclamation, [Reclamation] and U.S. Army Corps of Engineers), state agencies (e.g. State agencies (for example, California, Colorado), local agencies, and private enterprises.

The data typically used for flood frequency analysis consist of annual peak discharge values or peak discharges above a base value (~~partial-duration series~~PDS). Most annual peak records are obtained either from a continuous trace of river stages or from periodic observations provided by a crest-stage gage (~~Figure~~CSG, fig. 1). Crest-stage records may provide information only on peaks above some ~~pre-selected~~preselected base. The records are usually continuous, although missing data or zero flow years may be present. A statistical analysis of these data is the primary basis for the determination of the flow-frequency curve for each station. A major portion of these data are available in the ~~U.S. Geological Survey~~USGS National Water Information System (NWIS) and other electronic files; additional information in published or unpublished form is available from many sources (~~Appendix~~appendix 3).

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.





**Figure 1.** Photograph of a streamflow-gaging station showing a water-stage recorder, sharp-crested weir and crest-stage gage at U.S. Geological Survey station 01589238, Gwynns Falls Tributary at McDonogh, Maryland.

## Historical Flood Information

At many locations, particularly where people have occupied the flood plain for an extended period, or where civil works projects have been constructed by Federal agencies, there is information about major floods which occurred either before or after the period of systematic data collection. Similar information may be available at sites where the gage has been discontinued, or where records are broken or incomplete. Data for recent floods that occurred outside the systematic data collection period are also treated as historical floods. This historical flood information can often be used to make estimates of peak discharge. It also may define an extended period during which the largest floods, either recorded or historic, are known. In many cases, people make a physical mark, that which represents the approximate high-water mark of



**Historic-**

**Figure 2.** Photograph of historic flood high-water marks and flood of March 13-15, 2010, Potomac River at Great Falls Park, Virginia, upstream of U.S. Geological Survey streamflow-gaging station 01646500, Potomac River near Washington, D.C.

a flood (Koenig and others, 2016), on a relatively permanent surface (Figure fig. 2). The high-water mark elevation must be tied to a known datum in order to determine the peak discharge from a stage-discharge relation established after the flood.

Historical data are valuable information that are used in frequency analysis as follows: Let  $n_s$  denote the number of years in the systematic (gage) record,  $n_h$  be the number of years in the historical period and  $n$  be the total period of record, where  $n_s + n_h = n$ . Let  $T_h$  represent a discharge perception threshold that describes the knowledge that flood magnitudes exceeded this level, or were less than this level, during the historical period (Figure fig. 3). The historical flood data are represented by the historic ( $e_h$ ) peaks and the systematic ( $e_s$ ) peaks that exceed the threshold  $T_h$  during the total flood period  $n$ . There is also knowledge that during the historical period  $n_h$ , there are many years that no flood exceeded  $T_h$  (indicated with grey shading in Figure 3). The total number of floods that exceed the perception threshold is  $k$ , where  $k = e_s + e_h$ . The Section section Data Representation using Flow Intervals and Perception Thresholds discusses the determination of the historical period  $n_h$

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

## 8 Guidelines for Determining Flood Flow Frequency – Bulletin 17C

and estimation of perception threshold(s)  $T_h$ .

Historical data for flood frequency typically consists of ~~three types, the following three types~~ that can extend the temporal information on flood magnitudes:

- Large flood estimates prior to (outside of) the gaging station record (fig. 4);
- ~~an~~ An extraordinary large flood and knowledge that one (or more) floods within the gaging record are actually the largest in a longer time period  $n$  than that of the gaging station record  $n_s$  (~~Figure fig.~~ Figure fig. 5); and
- ~~knowledge~~ Knowledge that floods did not exceed some value  $T_h$  (~~non-exceedance~~ nonexceedance information) over a longer time  $n_h$  (~~Figure fig.~~ Figure fig. 6).

An example is used to illustrate each situation. In the first case, there are three historical floods that occurred prior to the establishment of the gaging station record. It is known that these floods exceeded a perception threshold of 18,000 cubic feet per second ( $\text{ft}^3/\text{s}$ ). These three floods are the largest on record, extend the observational record by 35 years (~~1895–1929~~ 1800s and 1850s), and are the most important for estimating flood frequency (~~Figure fig.~~ Figure fig. 4). In the second case, there is one extraordinary flood that occurred in June ~~1965~~ (Matthai, 1969, p. B39). This extraordinary flood is the largest in the 48-year gaging record (1948–1989), and there is historical flood and paleoflood information that indicates this flood might be the largest in over 900 years (Osterkamp and Costa, 1987) (~~Figure fig.~~ Figure fig. 5), rather than the largest in 48 years. Additional discussion for this extraordinary flood situation is in the ~~Section section~~ Extraordinary Floods. In the third case, one has information from a physical feature, such as a bridge or river terrace, ~~that where~~ no floods have exceeded a perception threshold. From detailed investigation of river terraces along the North Platte River near Seminoe dam, floods have not exceeded 45,000  $\text{ft}^3/\text{s}$  in the past 7,000 years (Levish, 2002; Levish and others, 2003) (~~Figure fig.~~ Figure fig. 6). Additional discussion for this situation is in the ~~Section section~~ Paleoflood and Botanical Information.

The USGS includes some historical flood information in its published reports and online. Addi-

tional information may be obtained from the files of other agencies, extracted from newspaper files, or obtained by intensive inquiry and investigation near the site for which the flood frequency information is needed (Thomson and others, 1964). Reports prepared by Federal agencies (such as the U.S. Army Corps of Engineers and ~~Bureau of~~ Bureau of Reclamation) to Congress requesting funding for civil works projects often contain historical flood information that supports the need for the project. These reports are available at many university and public libraries around the country. Data sources that could be used to identify the historical period  $n_h$ , perception threshold(s)  $T_h$ , and the largest floods outside the gaging record are described in ~~Appendix appendix~~ 3.

Over the past several decades, historical data and information have been shown to be extremely valuable in flood frequency analysis (Leese, 1973; Condie and Lee, 1982; Stedinger and Cohn, 1986, 1987; Cohn and others, 1997; England and others, 2003a). Dalrymple (1960) notes ~~“historical the following: “Historical~~ floods provide probably the most effective data available on which to base flood frequency determination, and where the data are reliable, this information should be given the greatest weight in constructing the flood frequency graph.” Historical flood information should be obtained and documented whenever possible. Use of historical data assures that estimates fit community experience and improves the frequency determinations. This information is valuable in flood frequency analysis because it directly contributes extreme flood data on low ~~annual-exceedance~~ annual-exceedance probability floods.

### Paleoflood and Botanical Information

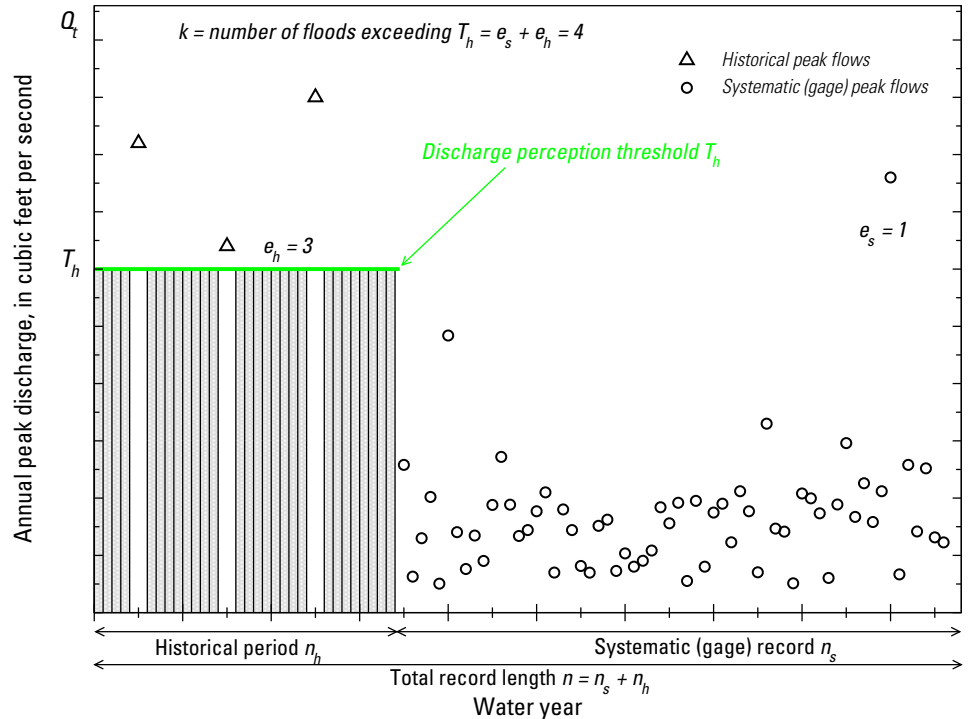
Over the past 40 years, there have been significant developments and advances in paleoflood hydrology (Costa, 1978, 1987; Baker and others, 1988; Jarrett, 1991; House and others, 2002a) ~~and~~ increased use of paleoflood data in flood frequency studies by Federal agencies and many others (Jarrett and Tomlinson, 2000; Levish and others, 2003; Sutley and others, 2008; Harden and others, 2011; O’Connor and others, 2014). Paleoflood hydrology primarily involves the study of floods that occurred before human record.

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017



Example peak-discharge

**Figure 3.** Graph showing an example peak-discharge time series with a historical period and a discharge perception threshold  $T_h$ . The grey shaded area represents floods of unknown magnitude less than  $T_h$  during the historical period  $n_h$ . Black vertical bars during the historical period represent flow intervals for each year for the unrecorded observations. Perception threshold  $T_h$  is shown as a green line. Historical floods that exceed the perception threshold (three 3 years) are shown as black triangles. Systematic (gage) peak flows are shown as black circles.

Paleofloods are different from historical floods in that they are determined by geologic and physical evidence of past floods rather than records based on community memory or referenced by built infrastructure. Paleoflood hydrology focuses on direct evidence of large, rare floods or the absence of such records. This is critical information for estimating the frequency of such floods (Baker, 1987; Baker and others, 2002).

Extraordinarily large floods often create geomorphologically significant changes to floodplains-flood plains and terraces, and leave evidence of flood stages in the geologic record that are long-lasting in time. Paleoflood data that are relevant for flood frequency typically consist of the following: paleostage indicators (PSIs) —(discrete evidence of maximum flood stages; and non-exceedance bound information —) and nonexceedance bound information (time intervals during which particular discharges have not been exceeded) (Levish, 2002) (see Glossary for complete

definitions). Paleoflood features that are typically used as PSIs for flood frequency are shown in Figure 7 (Jarrett and England, 2002); figure 7 (Jarrett and England, 2002) consist of slackwater deposits (SWDs) (Kite and others, 2002; House and others, 2002b); cobble and gravel flood bars (FBs) (Jarrett and England, 2002); tree scars (Yanosky and Jarrett, 2002); erosional scars and scour lines (Jarrett and Malde, 1987); and silt lines (O'Connor and others, 1986; Koenig and others, 2016). Geomorphic surfaces (primarily terraces) adjacent to rivers are used to place limits on flood discharges to estimate non-exceedance nonexceedance bounds (Levish, 2002).

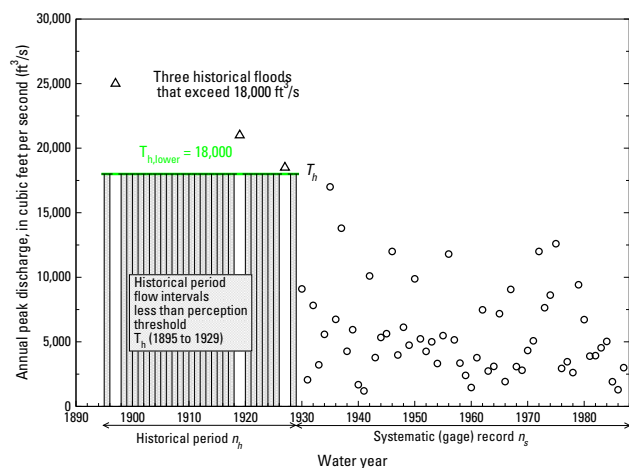
Paleoflood data-collection-data-collection methods and applications, including comprehensive overviews and current state of knowledge, are described in Baker and others (1988) and House and others (2002a). In many cases, paleoflood evidence persists for hundreds to thousands of years. This allows flood hydrologists

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

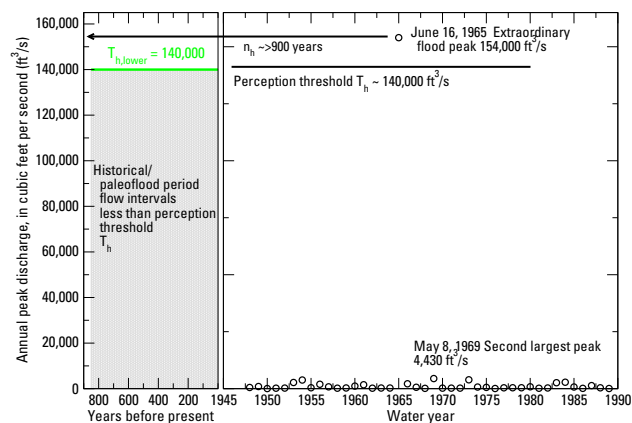
IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.





### Example

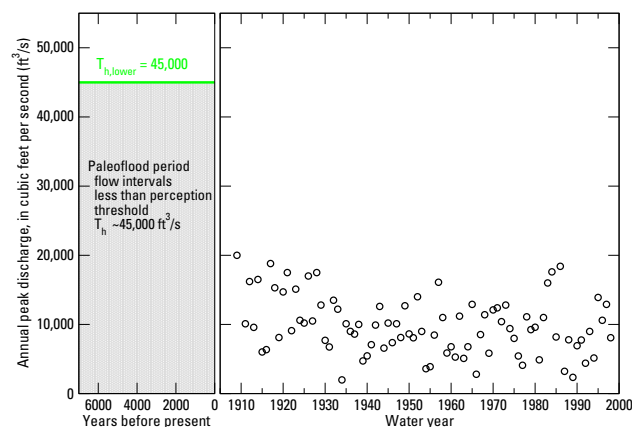
**Figure 4.** Graph showing an example site with three large historical floods outside the gaging record, Big Sandy River at Bruceton, Tennessee, U.S. Geological Survey streamflow-gaging station 03606500. The historical floods are known to exceed the perception threshold  $T_h$ .



### Example

**Figure 5.** Graph showing an example site with an extraordinary flood peak that represents a longer time frame, Plum Creek near Louviers, Colorado, U.S. Geological Survey streamflow-gaging station 06709500. A scale break is used to separate the gaging station data from the longer historical/paleoflood period. Horizontal lines indicate the approximate historical period  $n_h$  and the perception threshold  $T_h$ .

- 1 and geologists to obtain a great deal of relevant data
- 2 and information about the largest floods that have
- 3 occurred during an extended time period. Applied



### Example

**Figure 6.** Graph showing an example site with historical/paleoflood non-exceedance information, North Platte River near Seminoe Reservoir, Wyoming (Levish and others, 2003). A scale break is used to separate the gaging station data from the longer historical/paleoflood period. Floods have not exceeded a perception threshold  $T_h$  of 45,000 ft³/s in the past 7,000 years along the river; the largest floods in the gage record are 20,000 ft³/s.

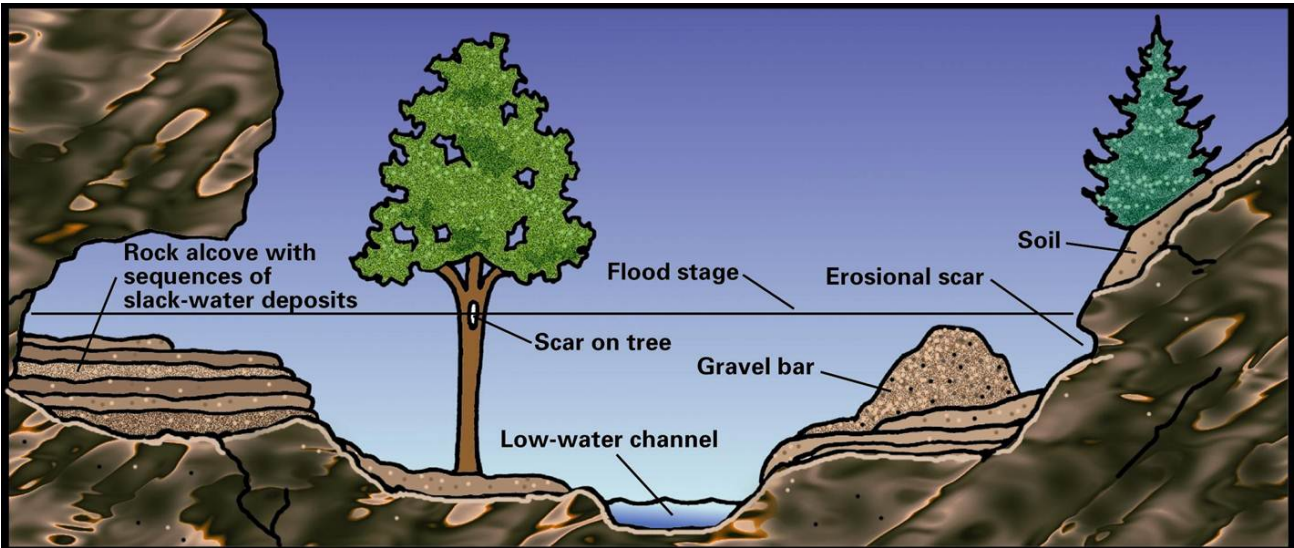
paleoflood and flood frequency studies (England and others, 2006; Harden and others, 2011) have shown that such evidence can greatly increase the precision of flood frequency estimates at a relatively low cost. In addition, these data are available now; one does not have to wait decades to obtain a substantially longer record.

Paleoflood data are treated in the same way as historical flood data for flood flow frequency analysis using these guidelines. Discharge perception thresholds for individual paleoflood magnitudes and non-exceedance nonexceedance bounds are used with age ranges for various paleoflood periods. In some cases, a single perception threshold, shown in Figure figure 3, is generalized to multiple thresholds for more complex paleoflood datasets (see Data Representation using Flow Intervals and Perception Thresholds Sectionsection). Paleoflood information should be obtained and documented whenever possible, particularly where the systematic record is relatively short and/or and (or) the AEPs of interest are small ( $\leq 0.01$ ). Some sources for paleoflood data, including regional approaches, are listed in Appendix appendix 3.

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.



Diagrammatic

**Figure 7.** [Diagram of a](#) section showing typical paleoflood features used as paleostage indicators (from Jarrett and England, 2002).

1 Botanical information consists of vegetation that  
2 records evidence of a flood (or several floods) and  
3 ~~for~~[\(or\)](#) indicates stability of a geomorphic surface for  
4 some time period. The types of botanical evidence  
5 utilized in paleohydrology studies consist primarily of  
6 age investigations, placement, distribution, and damage  
7 to trees. The four major types of botanical evidence  
8 of floods are (Hupp, 1987): corrasion scars, adventitious  
9 sprouts, tree age, and ring anomalies. Scars are the most  
10 easily observed damage feature, although outward evidence  
11 may disappear after a few years.

13 Sprouts generally occur from broken or inclined  
14 tree stems, sometimes called “clipper ships” ([Figure](#)  
15 [fig. 8](#)). Tree age may be utilized to date a particular  
16 flood or a geomorphic surface that has been inundated  
17 by a flood or may indicate the relative stability of a  
18 surface. Vegetation ages in both cases represent a  
19 minimum age since the surface was created. In some  
20 cases, trees trunks may be partially buried by flood-  
21 transported sediments; tree ages in this case are older  
22 than the geomorphic surface. Different tree ring patterns  
23 (eccentric, shifts, vessel changes, ~~etc.~~[etcetera](#))  
24 occur due to floods. Currently, the most reliable  
25 and accurate method of tree-ring-determined dates of

26 flooding is the analysis of increment cores or cross  
27 sections through scars (Hupp, 1988). Annual formation  
28 of rings permits flood dating to within a year, and  
29 sometimes to within several weeks ~~Yanosky and Jarrett (2002)~~ [\(Yan](#)  
30 Detailed descriptions of each type of evidence are pre-  
31 sented in Sigafos (1964), ~~Yanosky (1983)~~, ~~Hupp (1987, 1988)~~,  
32 [; Yanosky \(1983\); Hupp \(1987, 1988\);](#) and Yanosky  
33 and Jarrett (2002). Hupp and Osterkamp (1996)  
34 review the role of vegetation in fluvial geomorphic  
35 processes, including extreme floods. In flood frequency  
36 analysis, it is common to describe botanical  
37 information as binomial-censored observations corresponding  
38 to exceedances of a perception threshold. Some sources  
39 for botanical data are listed in [Appendix](#)  
40 [appendix 3](#).

### Common Issues with At-Site Data Records

42 There are several common issues associated with  
43 streamflow data records from gaging stations that may  
44 require investigation and treatment by the analyst.  
45 These issues include handling of incomplete records,  
46 extraordinary floods, and potentially ~~influential~~ [influential](#)  
47 low floods (PILEs).

PROVISIONAL---  
THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.  
IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).  
IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.



~~Inclined~~

**Figure 8.** ~~Photograph of inclined~~ western juniper trees with upright branches from ~~1861–1862~~ ~~1861–1862~~ flood, Crooked River near Prineville, Oregon.

## Broken, Incomplete, and Discontinued Records

Annual peaks for certain years may be missing because of conditions not related to flood magnitude, such as gage removal. These records are considered “broken”; a typical example is shown in ~~Figure-figure~~ 9. In this case, the analyst needs to determine if the records are equivalent, and if there is additional information such as historical or paleoflood information that can place the largest floods in a longer time context (~~Paretti and others, 2014a, Figure 4~~) (~~Paretti and others, 2014a, fig. 4~~).

The different record segments can be analyzed as a continuous record with length equal to the sum of both records if the gage is reestablished in a nearby location, unless there is some physical change in the watershed between segments, which may make the total record ~~non-homogeneous~~ ~~nonhomogeneous~~. Data from an upstream or downstream gage may provide additional information to estimate a perception threshold on the magnitude of floods that occurred during the missing or broken period.

An “incomplete” record refers to a streamflow record in which some peak flows are missing because they were too low or too high to record, or the gage was out of operation for a short period because of flooding. Missing high and low data ~~require-requires~~ different treatment. When one or more high annual

peaks during the period of systematic record have not been recorded, there is usually information available from which the peak discharge can be estimated, or a flow interval estimate can be made. A perception threshold is used to describe the knowledge that floods are not measured above a certain stage. For example, the ~~USGS National Water Information System (NWIS)~~ ~~NWIS~~ provides a code “8” that ~~indicates~~ a discharge was greater than an indicated value. At some crest gage sites, the bottom of the gage is not reached in some years. ~~the-USGS-The~~ ~~NWIS~~ provides a code “4” that a discharge was less than an indicated value. For this situation, a perception threshold is set to properly represent the incomplete observations less than some value. In most instances, the ~~data-collecting~~ ~~data-collecting~~ agency provides information to estimate peak discharges, flow intervals, and ~~for(or)~~ perception thresholds. Estimates that are made as part of the flood frequency analysis should be documented.

Streamflow-gaging data are available at many locations where records are no longer being collected. These stations and records are considered “discontinued”; they are extremely valuable, and should be utilized for frequency analysis. Streamflow records in many watersheds have been discontinued due to watershed development, including construction of dams and reservoirs. These discontinued records can be extended with the use of reservoir records (~~Appendix~~ ~~appendix~~ 3) and a perception threshold (~~Figure-fig.~~ ~~fig.~~).

## Extraordinary Floods

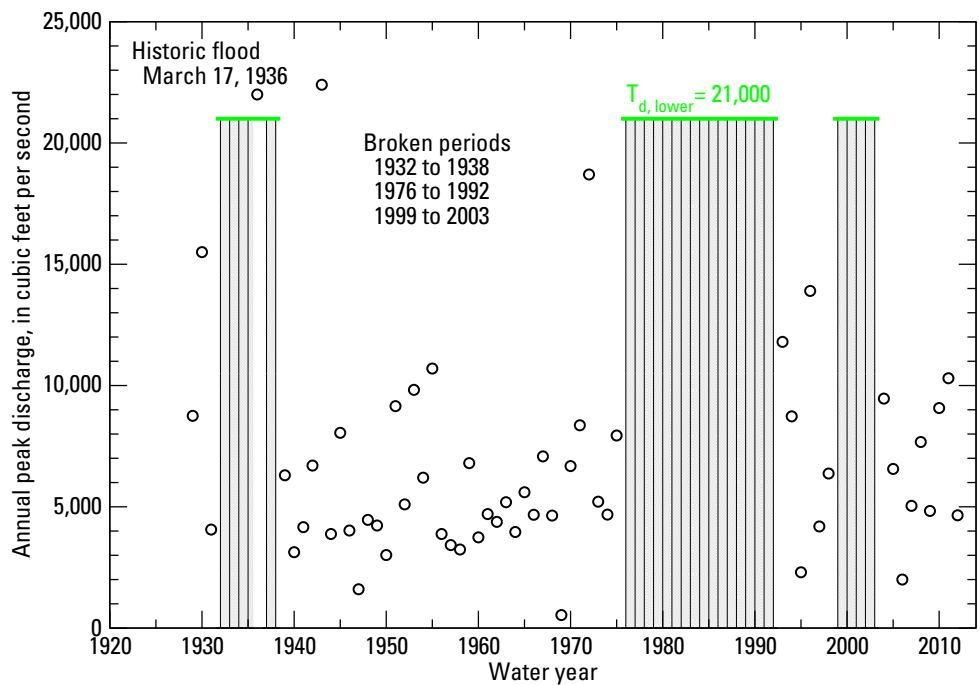
Extraordinary floods are those floods that are the largest magnitude at a gaging station or miscellaneous site and that substantially exceed the other flood observations (Costa and Jarrett, 2008). Extraordinary floods may be from gaging station records, indirect measurements at miscellaneous sites or from historical flood, paleoflood, or botanical information, as described in the ~~Sections-sections~~ Historical Flood Information and Paleoflood and Botanical Information. These floods typically exceed the second largest observation at a gaging station by a factor of two or greater, and in some cases, can be 35 times larger (~~Figure-fig.~~ ~~fig.~~ 5). There are many examples of extraor-

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.





Example-

**Figure 9.** Graph showing an example streamflow-gaging station with a broken record, U.S. Geological Survey station 01614000, Back Creek near Jones Springs, West Virginia. The grey shaded areas represent floods of unknown magnitude less than a perception threshold  $T_d$  (shown as a green line) during the systematic record  $n_s$  when the gage was discontinued. Black vertical bars during the systematic record represent flow intervals for each year for the unrecorded observations, with the perception threshold  $T_d$  based on the March 1936 flood.

dinary floods throughout the United States, such as the June 1921 flood on the Arkansas River in Colorado (Hazen, 1930) (Figure 10), fig. 10); the record 1954 flood on the Pecos River in Texas (Kochel and others, 1982; Lane, 1987); the 1976 Big Thompson River flash flood in Colorado (Costa, 1978; Jarrett and Costa, 1988); and the June 2008 Cedar River, Iowa flood (Eash, 2010). Costa and Baker (1981) describe some extraordinary floods that represent substantially longer time frames than the gaging record length  $n_s$  at each site. Costa and Jarrett (2008) discuss the physical process recognition and indirect discharge issues in estimating extraordinary flood magnitudes, and note that the uncertainty of these estimates is large.

These extraordinary floods are of critical importance because these estimates have a direct and large influence on the fitting of the flood frequency distribution, and are the events of interest to estimate flood magnitude and frequency. Extraordinary floods should be identified by using flood-peak-ratios, time

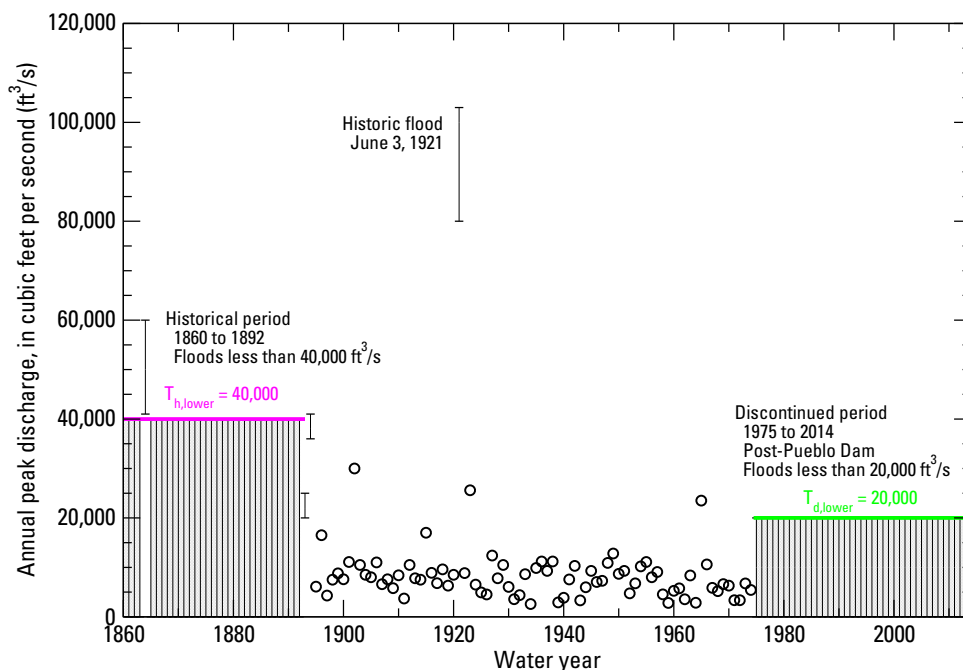
series-flood-peak ratios, time-series plots, and regional flood-peak-flood-peak envelope curves (Crippen and Bue, 1977; Asquith and Slade, 1995; O'Connor and Costa, 2004). The method used to estimate the extraordinary flood magnitude and relevant documentation should be reviewed to examine for potential errors, gather additional information about the flood, and to estimate uncertainty (Costa and Jarrett, 2008).

All extraordinary flood observations are to be retained and used in frequency analysis. These record floods represent a longer time-frame timeframe than that of the gaging record length  $n_s$ . Historical flood, paleoflood, and botanical information should be collected within the watershed and region of interest, in order to estimate perception thresholds  $T_h$  and expand the record length  $n_h$  for the extraordinary flood(s). The recommended procedures described in the Section section Determination of the Flood Flow Frequency Curve are appropriate for analyzing extraordinary floods at gaging stations. The use

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.



Example

**Figure 10.** Graph showing an example streamflow-gaging station with a discontinued record that is extended with a perception threshold from reservoir records, U.S. Geological Survey station 07099500, Arkansas River near Pueblo, Colorado. The gage was discontinued in 1975; floods since then have not exceeded a 20,000 cubic feet per second ( $\text{ft}^3/\text{s}$ ) perception threshold  $T_{d,l}$  shown as a green line. Large floods that occurred in 1864, 1893, 1894, and the June 3, 1921 extraordinary flood are described as interval observations and are shown as vertical bars with caps that represent lower and upper flow estimates. A perception threshold  $T_h$  for the historical period is shown as a magenta line.

of other frequency distributions, estimation procedures, or more complex models for extraordinary floods is not warranted, unless there is a need to estimate extreme flood quantiles less than 0.01 AEP as described in the section Frequency Curve Extrapolation. It is recommended to closely examine the fit of the flood frequency curve to the largest observations, and understand the influence of the any extraordinary observations on the fitted frequency curve. Confidence intervals should be used to estimate the range of AEPs for the flood. Examination of and comparison with regional flood information is also warranted. Regional flood-peak-flood-peak envelope probabilities (Vogel and others, 2007) can be considered in order to assess frequency estimates.

## Zero Flows and Potentially-Influential Potentially Influential Low Floods

Many rivers and streams in arid and semi-arid regions within the western-Western United States, such as in California (Lamontagne and others, 2012) and Arizona (Paretti and others, 2014a), have zero or very small flows for the entire year. The annual flood series for these streams will have one or more low-magnitude or zero flood values (Figure fig. 11). Such observations merit special attention. In particular, the logarithm of zero is negative infinity, and the logarithm of unusually small values can also be anomalous. Moreover, small flood values can have a large influence on the fitting of the flood frequency distribution and the estimation of the magnitude of rare flood flows. These small observations are called Potentially-Influential Potentially Influential Low Floods (PILFs) (Cohn and others, 2013).

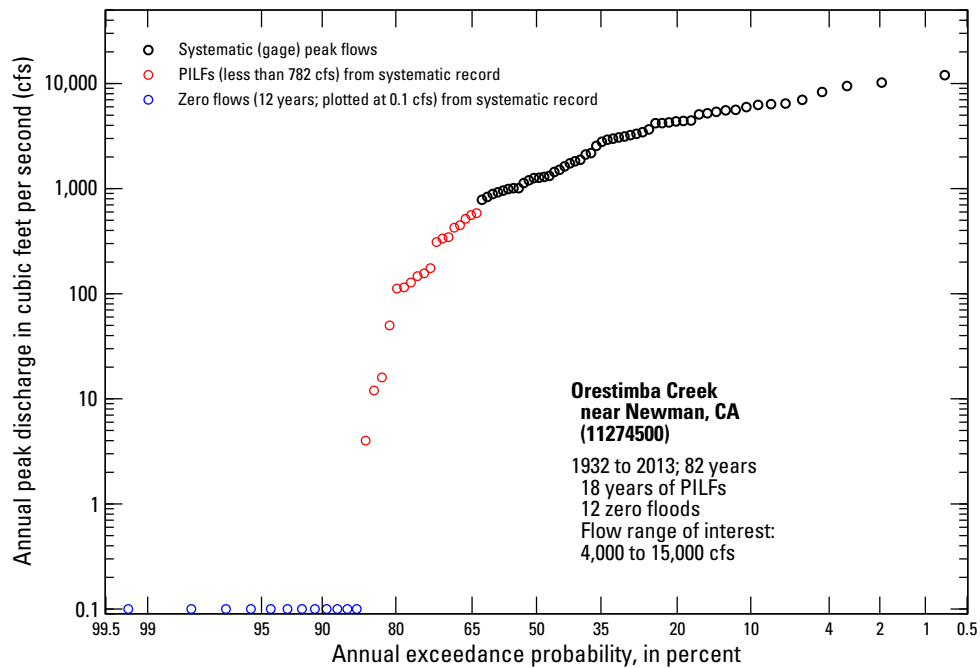
In these watersheds, the processes that create very

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.





Example-

**Figure 11.** Graph showing an example empirical frequency distribution and potentially ~~influential~~ low floods (PILFs) at U.S. Geological Survey station 11274500, Orestimba Creek near Newman, California.

large floods ~~i.e.~~ (that is, the floods of interest) may be different from the processes that cause the low (or zero) value annual peaks. Many low values can occur due to the influence of basin characteristics, such as channel-infiltration losses or evapotranspiration exceeding annual rainfall (Paretti and others, 2014a). The result is that the series of annual peaks appears to be generated from a mixed distribution. For example, peak flows in the range of 5,000 to 15,000 ft<sup>3</sup>/s are of interest on Orestimba Creek (Figure 11), rather than the numerous zero values and small flows less than about 1,000 ft<sup>3</sup>/s at this site. Consequently, the magnitudes of small annual peaks typically do not reveal much about the upper right-hand tail of the frequency distribution, and ~~thus~~, thus, should not have a highly influential role when estimating the probabilities of large floods (Cohn and others, 2013). These low (or zero) flows are ~~thus~~, thus, not relevant to estimating the probabilities of the largest flood events (Klemeš, 1986, 2000).

These Guidelines recommend the use of ~~robust~~ robust estimation procedures (Kuczera, 1982; Lamontagne and others, 2013) and a focus on the largest

floods ~~the~~ the upper tail of the flood frequency distribution (National Research Council, 1988) ~~to~~ to eliminate PILFs. Robust estimation procedures are reasonably efficient when the assumed characteristics of the flood distribution are true, while not doing poorly when those assumptions are violated (Stedinger and others, 1993; Cohn and others, 2013). A focus on the most extreme events (upper tail) is based on the observation that hydrometeorological and watershed processes during extreme events are likely to be quite different from those same processes during more common events (National Research Council, 1988, p. 7). The statistical procedure presented in the ~~Section~~ section Zeros and Identifying Potentially Influential Low Floods is used to detect PILFs.

### Data Representation using Flow Intervals and Perception Thresholds

Traditionally, flood flow frequency determination focused on the analysis of flood observations  $Q$  recorded in every year  $Y$  at continuous-record stream gages, which could be represented as point data  $Q_Y$ . The

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN. IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS). IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

description of flood and streamflow data for frequency analysis, and knowledge of the statistical characteristics of the data, have changed over the past 30 years. Valuable flood data ~~that~~ cannot usually be represented as point values ~~includes~~ include those from crest-stage gages, historical information, and paleoflood and botanical information. A generalized representation is used to capture what is known about annual peak flows in a given year  $Y$ , or over a range of years  $n$ . This includes information about specific annual floods that are known to be within a range of values, or above or below an estimated perception threshold. Also, there may be information over a range of years in which it is known that no flood occurred above a known perception threshold. There may be sites where multiple perception thresholds are needed to represent different segments of the sample data across the historical period.

Representations of peak-flow observations are now generalized to include concepts such as ~~flow~~ intervals, exceedances, nonexceedances, and multiple perception thresholds. These concepts are described in this section to provide a generalized data representation for flood frequency. Selected definitions and symbols for these concepts are presented in the Glossary. The recommended procedures in these Guidelines, described in the Section section Determination of the Flood Flow Frequency Curve, can readily incorporate these new types of information, and can use the data properly in frequency analysis of large floods. This allows use of all types of information in multiple combinations as necessary to best utilize the flood data available at a site. In these Guidelines, all flood data are represented by flow intervals and perception thresholds (Figure fig. 12).

For each year  $Y$ , the magnitude of  $Q_Y$  is characterized as a flow interval ( $Q_{Y,lower}, Q_{Y,upper}$ ). A lower estimate  $Q_{Y,lower}$  and upper estimate  $Q_{Y,upper}$  (interval) ~~is~~ are made based on observations, written records, or physical evidence. For the majority of floods, such as those from a gaging station, the discharge is nearly “exactly” known (for all practical purposes), and  $Q_{Y,lower} = Q_{Y,upper} = Q_Y$ . Floods that are described by intervals or ranges currently address the following two situations (Figure fig. 12): (1) a flood that is known to exceed some level, with no upper esti-

mate (binomial-censored data); and (2) floods that are known to fall within a large range (interval data). In the binomial case, one only knows the lower estimate  $Q_{Y,lower}$ ; the upper estimate  $Q_{Y,upper} \cong +\infty$  ~~and~~ is represented by a dashed line (Figure fig. 12). Flow intervals are used to describe, in some cases, the largest flood magnitudes that are estimated from historical and paleoflood records, and sometimes indirect measurements or field estimates at a gage ~~that~~ which have large uncertainty ( $\geq 25\%$ ). Flow intervals  $Q_{Y,lower}, Q_{Y,upper}$  are not used to provide ranges on gaged flows and reflect measurement uncertainties that are within ~~5-25~~ 5-25 25%. The interval observations are shown in Figure figure 12 with bars for the lower and upper estimates. For unobserved historical floods whose magnitudes are only known to be less than some perception threshold ( $T_h$ ), the lower estimate  ~~$Q_{Y,lower} = 0$~~   $Q_{Y,lower} = 0$ , and the upper estimate  $Q_{Y,upper}$  corresponds to the perception threshold for that year, such as  $T_{h1,lower}$  or  $T_{h2,lower}$  (Figure fig. 12). For crest-stage gages, flow intervals are determined with consideration of equipment recording limits of stage. There is usually a base (minimum) discharge  $Q_b$  established; this may vary each year.

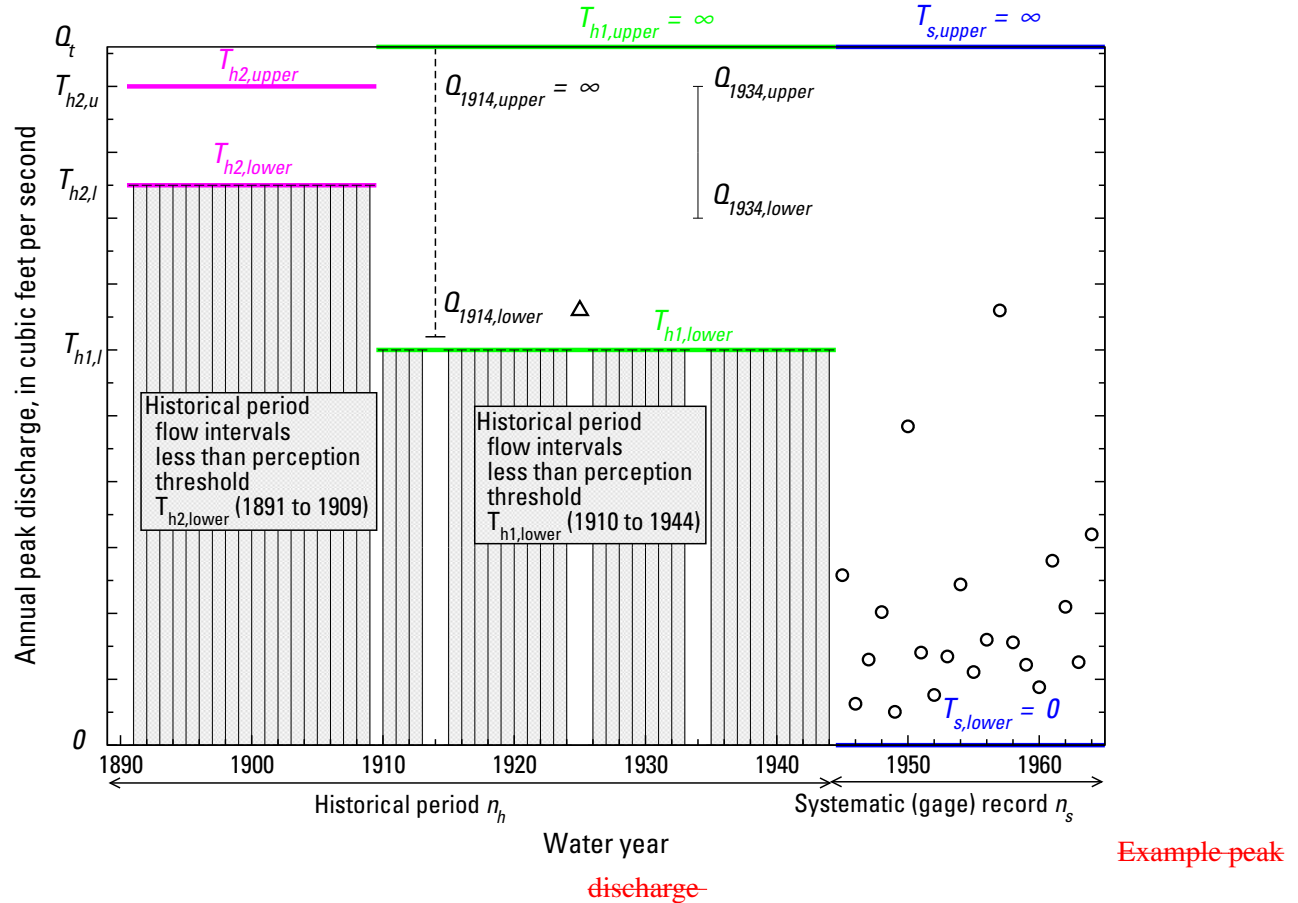
Perception thresholds ( $T_{Y,lower}, T_{Y,upper}$ ) are used to describe the knowledge in each year  $Y$  within the flood record, for which the value of  $Q_Y$  would have been observed or recorded. The lower bound ( $T_{Y,lower}$ ) represents the smallest peak flow that would result in a recorded flow; the upper bound ( $T_{Y,upper}$ ) represents the largest peak flow that could be observed or recorded. The interval ( $T_{Y,lower}, T_{Y,upper}$ ) defines the range of “perceptible values” ~~the~~ the range of potentially measurable flood discharges. These perception thresholds reflect the range of flows whose magnitude would have been recorded had they occurred, and are a function of the type of data collected at or near a gaging station and the physical characteristics of the river. In other words, the perception thresholds represent the “observable range” of floods. It is important to note that the perception thresholds  $T_Y$  do not depend on the actual peak discharges  $Q_Y$  that have occurred.

Lower and upper perception thresholds  $T_Y$  need to be estimated for each and every year of the record. The lower bound  $T_{Y,lower}$  represents the smallest annual peak flow that would result in a permanent record.

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.



**Figure 12.** Graph showing an example peak-discharge time series showing with peak flows, interval and binomial-censored flood observations, flow intervals, and perception thresholds. Systematic (gage) peak flows are shown as black circles, where  $Q_{Y,lower} = Q_{Y,upper}$ . During the historical period, there are the following three floods: a “binomial” observation in 1914; one flood with known magnitude in 1925; and an interval observation in 1934. The 1925 peak flow is shown as a black triangle, where the magnitude is exactly known ( $Q_{1925,lower} = Q_{1925,upper}$ ). The 1914 flood is described as a binomial observation and shown with a dashed line; it is known that this flood exceeded  $Q_{1914,lower}$  but the upper estimate is unknown. The flood in 1934 is known to fall within a certain range described with an interval ( $Q_{1934,lower} < Q_{1934,upper}$ ). Flood intervals are shown as black vertical bars with caps that represent lower and upper flow estimates. The grey-shaded areas represent floods of unknown magnitude less than the perception thresholds  $T_{h1,lower}$  and  $T_{h2,lower}$  during the historical period. The green lines represent the range in which floods would have been measured or recorded for the period 1910–1945, with lower and upper perception thresholds  $T_{h1,lower}$  and  $T_{h1,upper}$ . The magenta lines are the perception thresholds  $T_{h2,lower}$  and  $T_{h2,upper}$  for the period 1891–1909. The perceptible range for the systematic (gage) period (1945–1965)  $T_{s,lower}$ ,  $T_{s,upper}$  (0,  $\infty$ ) is shown as blue lines.

1 For systematic (gaging) records, this is typically rep-  
 2 resented by the “gage-base discharge,” which is typi-  
 3 cally 0. At crest-stage gages,  $Q_b > 0$ , and may  
 4 vary. For historical floods,  $T_{Y,lower}$  is typically esti-  
 5 mated to be equal to a historical flood discharge thresh-  
 6 old  $T_h$  (Figure fig. 12). For most sites with a system-  
 7 atic, continuous gaging record,  $T_{Y,upper}$  is assumed to

be infinite; larger floods typically get recorded. At  
 crest-stage gages, and for historical and paleoflood  
 periods,  $T_{Y,upper}$  needs to be estimated based on the  
 CSG recording range, historical information (such as  
 markers, bridges or buildings), or from geologic or  
 botanical evidence. For periods where the gage has  
 been discontinued (broken record) or ceased operation,

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
 PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
 PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
 GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
 TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

the observation thresholds are both set to infinity, if there is no other information such as a gage base or historical information. By setting  $T_{Y,lower} = T_{Y,upper} = \infty$ , this means that there is no information about that particular year. If there is historical information that is used for record extension of the largest floods during broken record periods,  $T_{Y,lower}$  can be set to a historical flood discharge threshold  $T_h$ . Some examples of flow intervals and perception thresholds are provided in [Appendix appendix 3](#).

In some situations, flood ~~data-sets~~ [datasets](#) need to be represented by multiple perception thresholds. This means more than one perception threshold is required to describe the data at hand. Two example perception thresholds are shown during the historical period in [Figure-figure 12](#). The period ~~1890-1909~~ [1890-1909](#) represents very limited observations prior to the establishment of a permanent settlement in a river canyon, where high stages and ~~HWMs~~ [high-water marks \(HWMs\)](#) that exceed  $T_{h2,upper}$  are not preserved or documented. The period ~~1910-1944~~ [1910-1944](#) represents increasing observations and awareness of all high river stages that exceed  $T_{h1,lower}$  (with no upper limit), prior to the establishment of a gaging station in 1945. It is appropriate to utilize multiple perception thresholds, particularly with longer historical records and paleoflood data, to properly represent the data and information at hand. In this situation, the two perception thresholds shown in [Figure-figure 12](#) would be extended with additional perception thresholds that are larger in magnitude than  $T_{h2,lower}$  and represent longer ~~time~~ [timeframes](#).

It is critical to collect historical data and determine the historical period  $n_h$  for flood frequency. The beginning of the historical period may be based on, for example, the earliest known historical settlement dates (such as 1860) along a river ([Figure fig. 10](#)), from archaeological information, or from paleoflood information and dating of river terraces and ~~non-exceedance bounds~~ ([Figures-nonexceedance bounds \(fig. 5 and 6\)](#)). The historical period does not begin at the earliest (first) observed flood, which is a biased estimate of  $n_h$ , as it is a lower bound on the true historical period (Hirsch and Stedinger, 1987).

The lower perception threshold  $T_{Y,lower}$  is particularly important to estimate. It represents our best

judgment, for any given year, of the smallest size flood that would have left evidence that the investigator would know about today. The historical or paleoflood information needs to persist so that hydrologists and geologists can obtain the data from written records, historical investigations, or paleoflood studies. For example, for every year during the period ~~1891-1909~~ [1891-1909](#), no evidence was found to indicate peak flows  $Q_Y$  had exceeded  $T_{h2,lower}$  ([Figure fig. 12](#)). The investigator should recognize that the lower limit of the perception threshold may be a rough approximation, and that it usually changes (increases in magnitude) as one moves backwards in time. In some cases, only the most catastrophic events would have been recorded and the threshold is high ([Figure fig. 5](#)); these are the events that are of interest.

## Regional Information and Nearby Sites

Flood information from within a region surrounding the gage site or watershed of interest is useful to improve flood frequency estimates, particularly when streamflow-gaging records are short (less than 30 years) (Stedinger and others, 1993). For these and other modest-length records, it is known that the station skew coefficient is sensitive to extreme events (Griffis and Stedinger, 2007a, 2009). Since Bulletin 17 (Beard, 1974), regional skew information  $G$  has been used to stabilize the station skew coefficient ( $\hat{\gamma}$ ), which defines the shape of the fitted frequency distribution, through the use of a “weighted” skew coefficient  $\tilde{G}$ . The techniques for estimating regional skew have evolved over the past 30 years (Tasker and Stedinger, 1986; Griffis and Stedinger, 2007d; Parrett and others, 2011), with the result that estimates are now much more accurate and their statistical properties are better understood, than at the time Bulletin 17B was written. It is recommended that regional skew information  $G$  is considered and weighted appropriately when estimating flood frequency curves. Some sources of regional skew information are listed in [Appendix-appendix 3](#). Additional guidance is provided in the [Sections-sections](#) Estimating Regional Skew and Weighted Skew Coefficient Estimator.

Other types of regional information that may be valuable for flood frequency ~~PROVISIONAL~~ can be considered, in

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.



addition to regional skew information. Griffiths and Stedinger (2007a) describe several flood frequency estimators and show that regional estimates of the mean and standard deviation can be valuable. In arid and semi-arid semiarid regions, regional mean and standard deviation estimates from peak flows can be used to improve at-site flood frequency estimates, such as the desert region in California (Gotvald and others, 2012). Physiographic characteristics within a watershed or region, such as mean basin elevation, drainage area, mean annual precipitation, and other physical factors, are useful in estimating regional parameters and in conducting regional flood frequency studies. Such studies usually employ generalized least-squares regression techniques (Tasker and Stedinger, 1989) to provide regional flood quantile estimates and quantile variances. These estimates are available for many states (Gotvald and others, 2012; Eash and others, 2013) States (Gotvald and others, 2012; Eash and others, 2013) and may be valuable for record extension and weighting of independent estimates. Additional guidance is provided in the Sections sections Record Extension with Nearby Sites and Weighting of Independent Frequency Estimates.

## Flood Estimates from Precipitation

Flood discharges estimated from climatic data (rainfall and ~~for~~<sup>or</sup> snowmelt) can be a useful addition to direct streamflow measurements. Estimates may be available from several cases, such as: (1) flood estimates from individual extreme events that are based on observed rainfall; (2) synthetic flood events and frequency curves from rainfall frequency estimates; and (3) continuous streamflow estimates and frequency curves from precipitation and climate information.

Such estimates require at least adequate climatic data and a valid watershed model for converting precipitation to discharge. In some situations, existing watershed models may be available that are already calibrated to the watershed of interest. For example, the National Weather Service (NWS) has calibrated watershed models for flood forecasting on major river basins through their River Forecast Centers (RFCs). Other Federal agencies (~~USACE~~<sup>U.S.</sup> Army Corps of Engineers, Reclamation, ~~NRCS~~<sup>and</sup>

~~Natural Resources Conservation Service~~) may have calibrated flood watershed models for flood control, levee design, and other projects within their jurisdiction. As part of ~~floodplain~~<sup>flood-plain</sup> management studies for the Federal Emergency Management Agency (FEMA), ~~state~~<sup>State</sup> agencies, counties, and local watershed protection districts may have calibrated watershed models for large floods that may be used to supplement streamflow-gaging station records.

Individual extreme floods or flood frequency curves can be estimated from event-based or continuous rainfall-runoff models (National Research Council, 1988; Singh, 1995; U.S. Bureau of Reclamation and Utah State University, 1999; Beven, 2001; FEMA, 2009), using observed watershed precipitation, precipitation observed at nearby stations in a meteorologically homogeneous region, or from stochastically ~~generated~~<sup>generated</sup> precipitation. The rainfall-runoff model needs to be calibrated to extreme flood observations, using procedures such as those presented in Duan and others (2003), in order to be useful for flood frequency estimation and prediction. It is recommended that an uncertainty analysis be conducted (Kjeldsen and others, 2014), including prediction uncertainty (Beven, 2001, Chapter 7), to reflect the range of variability associated with the estimated flood frequency curve from the rainfall-runoff model. The variance of flood quantile estimates from rainfall-runoff models is also needed for potential weighting of the estimate, as described in the ~~Section~~<sup>section</sup> Weighting of Independent Frequency Estimates.

Flood frequency estimates from rainfall-runoff models can be biased low (Thomas, 1982) or high and exhibit a loss of variance (Lichty and Liscum, 1978; Thomas, 1987) when model and other errors are not properly accounted for in uncertainty analysis. ~~Including~~<sup>The inclusion of</sup> variability in precipitation and temperature inputs (Clark and others, 2004) helps in this situation. In some cases, rainfall-runoff models are calibrated to or parameters are adjusted to better match flood frequency curves based on peak-flow statistics (Reed, 1999; Swain and others, 2006; MGS Engineering Consultants, 2009). Frequency curves from rainfall-runoff models need to be independent of the frequency curve estimated using the recommended procedures in these Guidelines, if curves are to be

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

weighted and combined.

Analysts making use of such procedures should clearly document the rainfall-runoff method used for computing the floods and evaluate its performance based upon flood and storm experience in a hydrologically and meteorologically homogeneous region, including calibration and uncertainty analysis. Whether or not such studies are useful will depend upon the availability of the information, the adequacy of the existing flood records, and the purpose for which the watershed model was developed and calibrated. The magnitude and AEP of the precipitation or flood event are the most important factors to consider when including these estimates. The largest or most extreme flood events, with  $AEPs < 0.02$  are very useful, especially for ungaged sites or in situations where gaging stations have been destroyed.

In addition to flood estimates from precipitation, hydroclimatological information (Maddox and others, 1980; Hirschboeck, 1991) is very useful and provides a broad perspective on data and flood processes for frequency analysis. Atmospheric circulation patterns (Hirschboeck, 1987a) and climate indices such as the El Niño-Southern Oscillation (ENSO) (Webb and Betancourt, 1992) can be coupled with streamflow records to gain insight to the types of flood-causing mechanisms and flood variability (National Research Council, 1999). Redmond and others (2002) describe important connections between climate mechanisms, paleofloods, and flood variability. Some sources for precipitation and climate information are listed in [Appendix 3](#).

## Data Assumptions and Specific Concerns

The conventional assumptions for a statistical analysis are that the array of flood information is [a reliable and reliable and it is a](#) representative time sample of random, homogeneous events. Assessment of the adequacy and applicability of flood records is [therefore, therefore,](#) a necessary first step in flood frequency analysis. This section discusses flow measurement error, randomness of events, mixed populations, watershed changes, and climate variability and change con-

siderations for flood frequency analysis.

## Flow Measurement Error

Peak-flow measurement errors exist in streamflow records, as in all other measured values. Sauer and Meyer (1992) [describes-describe](#) sources of error in streamflow measurement. Absolute errors in flow estimates are generally greatest during maximum flood flows. [Peak-flow-Peak-flow](#) estimates of the largest floods from systematic (gage) records, historical floods, paleofloods, or from other sources, can be substantially in error because of the uncertainty in both stage and stage-discharge relationships, and because the flows may be estimated from [rating-curve-rating-curve](#) extensions or indirect methods, rather than by direct measurement. Many improvements have been made in direct measurements of streamflow by the USGS over the past several decades (Turnipseed and Sauer, 2010), with “good” (5%) accuracy of most discharge measurements. However, the largest flows are generally not directly measured because of problems with debris, inaccessibility issues, and safety considerations (Costa and Jarrett, 2008). Other sources of potential error in large discharges include undocumented and unmetered breakout flows from the main river channel, and ice effects (Rantz and Others, 1982b). The largest floods are usually estimated by rating-curve extensions or indirect methods, with estimation errors that can exceed 25% in many cases, to over 100% in high-gradient streams (Jarrett, 1987). Measurement errors can seriously degrade flood quantile estimates in some situations (Potter and Walker, 1985); therefore, estimation errors in the largest floods should be investigated.

In many instances, annual peak discharges are estimated from rating-curve extensions. Significant errors in discharge estimation may occur from [rating-curve-rating-curve](#) extensions (Cook, 1987; Kuczera, 1996), especially if the discharge value is more than twice the greatest measurement by current meter. Unfortunately, high outliers or significant flood peaks are usually never measured directly and are many times greater than twice the measured value (Klemeš, 1987). Kuczera (1996) indicates that [rating-curve-rating-curve](#) extensions, in the presence of correlated errors, can

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

significantly affect quantile estimates from such extrapolations.

Indirect methods are utilized to measure peak discharges after flood periods (Benson and Dalrymple, 1967; Rantz and Others, 1982a), using ~~high-water marks (HWMs)~~ HWMs or PSIs (Koenig and others, 2016). The slope-area method (Dalrymple and Benson, 1967) is most commonly used by the ~~U.S. Geological Survey~~ USGS; other indirect methods are presented by Cook (1987) and Webb and Jarrett (2002). Slope-area methods have documented sources of uncertainty (Bathurst, 1986; Jarrett, 1987; Kirby, 1987; McCuen and Knight, 2006). Significant errors in indirect discharge estimates have been noted in mountain areas; the measurements typically are overestimated (Jarrett, 1987). Neglecting channel scour or fill is the most significant factor that may introduce large errors in indirect discharge estimates (Kirby, 1987). Quick (1991) presents sources of errors in the slope-area method and indicates that the method has a strong upward bias. Webb and Jarrett (2002) describe assumptions in estimating historical and paleoflood peak discharges; they also outline information needed to support discharge estimates.

At times errors will be apparent or suspected. Substantial efforts should be made to understand sources of flow measurement errors, and to quantify the uncertainty associated with such errors. If substantial, the errors should be brought to the attention of the data collecting agency, with supporting evidence and a request for a corrected value.

## Randomness of Events

In general, a time series of annual ~~peak-flow~~ peak-flow estimates may be considered to be a random sample of independent, identically ~~-distributed~~ distributed random variables. The peak-flow time series is assumed to be a representative sample of the population of future floods. This assumption is contingent upon conducting exploratory data analysis (~~Appendix~~ appendix 4) and further physical knowledge of the system. In essence, the stochastic process that generates floods is assumed to be stationary, or invariant in time. Stationarity is a property of an underlying stochastic process, and not of observed data.

Realizations from stationary processes can exhibit excursions and trends that persist for decades or centuries (Cohn and Lins, 2005). Nonstationary processes are difficult to detect in peak-flow series (Villarini and others, 2009a) and may be challenging to determine (Koutsoyiannis, 2011). In some situations, long-term persistence concepts (Lins and Cohn, 2011) or shifting-mean models (Salas and Boes, 1980; Sveinsson and others, 2003) could be considered.

Before conducting flood frequency analysis, these Guidelines recommend that analysts perform an initial analysis of the data. Helsel and Hirsch (1992) and Hirsch and others (1993) provide overviews and details on conducting exploratory data analysis. The recommended procedures for initial data analysis include plotting the series, estimating serial correlation, ~~and~~ examining for trends and abrupt shifts (change points), and are presented in ~~Appendix~~ appendix 4.

In certain locations, flood records may indicate apparent nonrandomness and exhibit strong multi-decadal trends or wet and dry cycles that are not explained by land use change, water management, or climate change. Such records are particularly challenging and this is one of the most vexing problems in flood frequency analysis. The Work Group did not evaluate methods to account for nonrandomness and ~~for~~(or) multidecadal trends in flood frequency. Additional work in this area is warranted, as it is a seriously unresolved problem. If multidecadal trends of this sort are identified through appropriate statistical tests and data analysis, it is recommended that the underlying physical mechanisms be investigated to gain hydrological understanding (Lins and Cohn, 2011). How to adjust such a record for flood frequency is an unresolved problem.

Even when statistical tests of the serial correlation coefficients indicate a significant deviation from the independence assumption, the annual peak data may define an unbiased estimation of future flood activity if other assumptions are attained. The ~~non-randomness~~ nonrandomness of the peak series will, however, result in error in the estimated uncertainty associated with the fitted frequency curve (Tasker, 1983). Effective ~~record-length~~ record-length concepts (Tasker, 1983; Vogel and Kroll, 1991) should be used to correct uncertainty estimates in the presence of serial correlation.

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

## Mixed Populations

Flooding in some watersheds is caused by different types of meteorological events associated with distinct physical processes. For example, flooding at some locations may arise from snowmelt, rainstorms, or by combinations of both snowmelt and rainstorms (Jarrett and Costa, 1988). Such a record may not be homogeneous and may require special treatment. This mixed population results in flood frequency curves with abnormally large skew coefficients reflected by abnormal slope changes when plotted on logarithmic normal probability paper. In some situations, the frequency curve of annual events can best be described by computing separate curves for each type of event ~~and~~ and then combining the results.

One example of mixed population is rainfall-runoff mixed with snowmelt. In the Sierra Nevada region of California, hydrologic factors and relationships operating during general winter rain floods are usually quite different from those operating during spring snowmelt floods or during local summer cloudburst floods. In this region, peak flows are primarily caused by winter rainfall at lower elevations, while at higher elevations, peak flows are generally caused by spring snowmelt or rain-on-snow events (Parrett and others, 2011). Frequency studies in the Sierra Nevada have been made separately for rain floods, which occur principally during the months of November through March, and for snowmelt floods, which occur during the months of April through July. Peak flows were segregated by cause ~~—~~ — those predominately caused by snowmelt and those predominately caused by rain (Crippen, 1978). Likewise, in the Colorado Front Range, peak ~~flows~~ flows are caused by both rainfall and snowmelt during the spring and summer (Elliott and others, 1982), especially in the lower elevation of the foothills zone (Jarrett and Costa, 1988).

Flooding in the eastern United States is caused by a mixture of flood-generating mechanisms, with tropical cyclones and extratropical systems playing a central role (Smith and others, 2011). Along the Atlantic and Gulf Coasts, in some instances, floods from hurricane and ~~non-hurricane~~ nonhurricane events have been separated, thereby improving frequency estimates (Murphy, 2001). Ice-jam floods that occur in

northern regions (Murphy, 2001) are another mixed-population example.

Hydroclimatological data, including the use of synoptic weather patterns (Hirschboeck, 1987b), is particularly useful to provide independent, physically ~~-based~~ -based information on climate-induced flood processes and to separate flood series by type. Additional data, such as paleohydrologic and paleoclimate data, may also be considered (Redmond and others, 2002). The flood types and particular causative mechanisms may also be explored using a watershed perspective and considering variables such as storm rainfall and duration, flood seasonality, timing, and runoff response (Merz and Blöschl, 2003).

When it can be shown that there are two or more distinct and generally independent causes of floods, it may be more reliable to segregate the flood data by cause, ~~analyze and compute~~ analyzing and computing separate curves for each type of event ~~and then to combine~~ and then combining the curves into an overall analysis of the flood frequency at the site. Procedures such as those described in Crippen (1978), U.S. Army Corps of Engineers (1982), Jarrett and Costa (1988), and Murphy (2001) may be considered. For ~~ice-jam~~ ice-jam flow situations, one may consider using the same mixed-population approach (Murphy, 2001), or a method that focuses on maximum elevation (Vogel and Stedinger, 1984). An example of combining frequency curves was performed for the Black Hills region as part of the ~~peak-flow~~ peak-flow frequency estimates for South Dakota (Sando and others, 2008) ~~;~~ ; Alila and Mtiraoui (2002) describe other examples. In some situations, there may not be sufficient data to perform a mixed-population analysis, or the results may not be as reliable (Gotvald and others, 2012). The Work Group did not conduct an evaluation of these procedures. Additional efforts are needed to provide guidance on the identification and treatment of mixed distributions.

Separation by calendar periods in lieu of separation by events is not considered hydrologically reasonable, unless the events in the separate periods are clearly caused by different hydrometeorological conditions. The fitting procedures in these Guidelines can be used to fit each flood series separately, with the exception that regional skew coefficients cannot be

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.



used unless developed for the specific types of events being examined. If the flood events that are believed to comprise two or more populations cannot be identified and separated by an objective and hydrologically meaningful criterion, the record shall be treated as coming from one population.

## Watershed Changes

It is becoming increasingly difficult to find watersheds in which the flow regime has not been altered by modifications to the river channel, to the river floodplain, creation or destruction of reservoirs and levees, or modifications to the characteristics of the watershed at large (e.g., for example, urbanization, wildfires, change of cropping practices, erosion control, land drainage, or deforestation). Developments which that can change flow conditions include urbanization, channelization, agricultural drainage, levees, and the construction of reservoirs, diversions, and alteration of land cover conditions (Sauer and others, 1983). Impervious areas within the watershed and their effects on runoff are also important considerations (Moglen, 2009).

Watershed history and flood records should be carefully examined to assure that no major watershed changes have occurred during the period of record. Documents which that accompany flood records often list such changes that occurred at discrete times. However, the effects of urbanization or the construction of numerous small reservoirs over a period of several years will likely not be documented. Such incremental changes may not noticeably alter the flow regime from year to year but the cumulative effect can be significant.

Special effort should be made to identify those records which that are not homogeneous. The data analysis tools described in Appendix 4 may be used to assess records for potential gradual trends or shifts that might be associated with watershed changes. Spatial and temporal estimates of land use data within watersheds should be obtained where they are available. These data are particularly useful in quantifying urbanization impacts on flood frequency (Moglen and Beighley, 2002; McCuen, 2003).

Only records which that represent relatively con-

stant watershed conditions should be used for frequency analysis (Konrad, 2003; Moglen and Shivers, 2006). In some situations, flow records may be adjusted to account for watershed change so that they represent current watershed conditions, where physical evidence of watershed change exist in a significant portion of the watershed (McCuen, 2003). The Work Group did not evaluate methods to account for watershed changes and makes no particular recommendations, as additional work is needed in this area.

## Climate Variability and Change

There is much concern about changes in flood risk associated with climate variability and long-term climate change. Time invariance was assumed in the development of this guide. In those situations where there is sufficient scientific evidence to facilitate quantification of the impact of climate variability or change in flood risk, this knowledge should be incorporated in flood frequency analysis by employing time-varying parameters or other appropriate techniques. All such methods employed need to be thoroughly documented and justified.

The Work Group did not evaluate methods to account for climate variability in flood frequency. Additional work in this area is warranted. Some information and background on nonstationarity is presented in Olsen and others (2010) and Kiang and others (2011). In the interim, analysts might consider the following:

- data on synoptic weather patterns (Hirschboeck, 1987b);
- paleoclimate information (Redmond and others, 2002);
- climate variability and climate projection information (Brekke and others, 2009);
- interannual and interdecadal variations in climate (Jain and Lall, 2001); and
- time-varying distribution parameters (Stedinger and Griffis, 2011; Salas and Obeysekera, 2014).

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

## Determination of the Flood Flow Frequency Curve

This section presents the recommended procedures for determining a flood flow frequency curve. The procedures include the following: approaches for plotting positions; the flood distribution; parameter estimation; methods to handle zeros and identifying PILFs; the ~~Expected Moments Algorithm~~EMA; record extension; and confidence intervals for quantiles. Computer programs are required in order to make these calculations; see the ~~Section~~section Software and Examples for available products.

### Plotting Positions

Empirical frequency distributions are a “nonparametric” or ~~distribution-free~~distribution-free method to infer the probability distribution function (mathematical model) that describes flood risk. They are used to assess distribution function (~~e.g.~~for example, LP-III) fits the data. Probability estimates are made using plotting positions. A basic plotting position formula is (Stedinger and others, 1993, p. 18.24) ÷

$$p_i = \frac{i - a}{n + 1 - 2a} \quad (2)$$

where  $p_i$  is the exceedance probability of flood observations  $Q_i$  ranked from largest ( $i = 1$ ) to smallest ( $i = n$ ), and  $a$  is a plotting position parameter ( $0 \leq a \leq 0.5$ ) (see ~~Table 5.1 in Appendix~~table 5.1 in appendix 5).

Historical flood peaks reflect the frequency of large floods and~~thus,~~thus, should be incorporated into flood frequency analysis. They can also be used to judge the adequacy of estimated flood frequency relationships. For this latter purpose, appropriate plotting positions or estimates of the average exceedance probabilities associated with the historical peaks and the remainder of the data are desired. Hirsch and Stedinger (1987) and Hirsch (1987) provide an algorithm for assigning plotting positions to censored data, such as historical floods. They emphasized the correct interpretation of the information conveyed by historical flood data, the recognition of the limited precision of estimates of the exceedance probabilities of historical floods, and showed that all

estimators were relatively imprecise (Hirsch and Stedinger, 1987). The ~~threshold-exceedance-plotting position~~threshold-exceedance plotting position formula is given in ~~Appendix~~appendix 5. It is applicable for ~~potentially-influential~~potentially influential low flood cases, in addition to historical data, as the censored-data principles are the same.

### Flood Distribution

Flood records describe a succession of natural events ~~which that~~ do not fit any one specific known statistical distribution. To make the problem of defining flood probabilities tractable, it is convenient to select a reasonable mathematical distribution. These Guidelines recommend the use of the log-Pearson Type III (LP-III) distribution. This distribution has been in use by Federal agencies since 1967 (USWRC, 1967; Benson, 1968).

Several studies have been conducted over the years to investigate which of many possible distributions and alternative parameter estimation procedures would best meet the purposes of these Guidelines. Beard (1974), summarized in IACWD (1982), found that the LP-III distribution with a regional skew coefficient performed well. Griffis and Stedinger (2007b) explored the the characteristics of the LP-III distribution ~~;~~and showed that it is very flexible and encompasses a wide range of reasonable models for log-space skews, such as  $|\gamma| \leq 1.414$ . The method of moments parameter estimation procedure works well with reasonable constraints on parameters (Griffis and Stedinger, 2007c) ~~;~~and an informative regional skew (Griffis and Stedinger, 2009). The Work Group concluded from these studies, many applications over the past 40 years, and testing (Cohn and others, 2017) ~~;~~ that the Pearson Type III distribution with log transformation of the data (log-Pearson Type III distribution) with a regional skew coefficient is the base method for analysis of annual peak-flow data. The LP-III distribution also performs well and is appropriate for applications with historical and paleoflood data (England, 1998; Bureau of Reclamation, 2002; Blainey and others, 2002; England and others, 2003a, 2010; Harden and others, 2011).

The base 10 ~~logarithms~~logarithms  $X_1 \dots X_n$  of peak flows

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

$Q_1 \dots Q_n$  are assumed to follow a Pearson Type III (P-III) distribution; this probability density function  $f(x)$  is :-

$$f(x|\tau, \alpha, \beta) = \frac{\left(\frac{x-\tau}{\beta}\right)^{\alpha-1} \exp\left(-\frac{x-\tau}{\beta}\right)}{|\beta|\Gamma(\alpha)}, \quad (3)$$

with  $\left(\frac{x-\tau}{\beta}\right) \geq 0$  and distribution parameters  $\tau$ ,  $\alpha$  and  $\beta$ , where  $\tau$  is the location parameter,  $\alpha$  is the shape parameter,  $\beta$  is the scale parameter, and  $\Gamma(\alpha)$  is the gamma function, defined as :-

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} \exp(-t) dt. \quad (4)$$

The shape parameter  $\alpha$  is limited to positive values, and the scale parameter  $\beta$  may be positive or negative. When  $\beta > 0$ , the P-III distribution has a lower bound  $\tau$  and is ~~positive-skewed~~ positive skewed; the distribution is negative skewed when  $\beta < 0$  ( $\tau$  is an upper bound). This behavior may also be described using the skewness coefficient, rather than parameters. When the skewness coefficient  $\gamma$  is greater than zero ( $\gamma > 0, \beta > 0$ ), the distribution has a positive skew and floods are unbounded. When  $\gamma < 0, (\beta < 0)$ , the distribution of the logarithm of floods has a negative skew and an upper bound. In this situation, the log-space skew is constrained to  $\gamma \geq -1.41$ . Griffis and Stedinger (2007b) present additional properties of the P-III distribution, including plots of the P-III probability density function.

## Parameter Estimation: Simple Case

These Guidelines recommend the method of moments using the logarithms of flood flows to estimate the parameters of the P-III distribution. The first three sample moments are used to estimate the P-III parameters. These include the mean ( $\hat{\mu}$ ), standard deviation ( $\hat{\sigma}$ ), and skewness coefficient ( $\hat{\gamma}$ ).

## Moments and Parameters

In the case where only systematic data are available, with no historical information or PILFs, the mean, standard deviation, and skewness coefficient of station data may be computed using the following

equations:

$$\hat{\mu} = \left(\frac{1}{n}\right) \sum_{i=1}^n X_i, \quad (5)$$

$$\hat{\sigma} = \sqrt{\left(\frac{1}{n-1}\right) \sum_{i=1}^n (X_i - \hat{\mu})^2}, \quad \text{and} \quad (6)$$

$$\hat{\gamma} = \left(\frac{n}{\hat{\sigma}^3(n-1)(n-2)}\right) \sum_{i=1}^n (X_i - \hat{\mu})^3 \quad (7)$$

where  $n$  is the number of flood observations and  $(\hat{\cdot})$  represents a sample estimate. The standard deviation ( $\hat{\sigma}$ ) and skewness coefficient ( $\hat{\gamma}$ ) include bias correction factors  $(n-1)$  and  $(n-1)(n-2)$  for small samples.

The parameters are estimated from the sample moments as :-

$$\hat{\alpha} = \frac{4}{\hat{\gamma}^2}, \quad (8)$$

$$\hat{\beta} = \text{sign}(\hat{\gamma}) \left(\frac{\hat{\sigma}^2}{\hat{\alpha}}\right)^{1/2}, \quad (9)$$

and

$$\hat{\tau} = \hat{\mu} - \hat{\alpha}\hat{\beta}. \quad (10)$$

Flood quantiles  $\hat{Q}_q$  for the P-III distribution can be estimated by

$$\hat{X}_q = \hat{\tau} + \hat{\beta}P^{-1}(\hat{\alpha}, q) \quad (11)$$

where  $P^{-1}(\hat{\alpha}, q)$  is the inverse of the incomplete Gamma function (Abramowitz and Stegun, 1964) and

$$\hat{Q}_q = 10^{\hat{X}_q} \quad (12)$$

where  $q$  is the cumulative probability of interest (~~e.g.~~ for example,  $q = 0.99$ , and  $q = 1 - p$ ). Flood quantiles can also be estimated with the use of a frequency factor  $K_{\hat{\gamma}, p}$  that is a function of the skew coefficient  $\hat{\gamma}$  and exceedance probability  $p$  ~~as shown by the following equation~~

$$\hat{X}_p = \hat{\mu} + \hat{\sigma}K_{\hat{\gamma}, p} \quad (13)$$

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

1 and  $K_{\hat{\gamma},p}$  can be estimated from available algorithms  
 2 (Kirby, 1972; Stedinger and others, 1993).

### 3 Weighted Skew Coefficient Estimator

4 There is relatively large uncertainty in the at-  
 5 site sample skewness coefficient (third moment)  $\hat{\gamma}$  ;  
 6 because it is sensitive to extreme events in ~~modest~~  
 7 ~~length-modest-length~~ records (Griffis and Stedinger,  
 8 2007a). The station skew coefficient  $\hat{\gamma}$  and regional  
 9 skew coefficient  $G$  can be combined to form a better  
 10 estimate of skew  $\tilde{G}$  for a given watershed, as illus-  
 11 trated by the concepts in Tasker (1978). Under the  
 12 assumption that the regional skew coefficient  $G$  is  
 13 unbiased and independent of the station skew  $\hat{\gamma}$ , the  
 14 ~~mean-square-mean square~~ errors (MSEs) of the the  
 15 station skew  $MSE_{\hat{\gamma}}$  and the regional skew  $MSE_G$  can  
 16 be used to estimate a weighted skew coefficient, as  
 17 described in ~~Appendix-appendix~~ 7. The MSE of the  
 18 station skew is computed directly by EMA. The MSE  
 19 of the regional skew is usually estimated through the  
 20 procedures described in the ~~Section-section~~ Estimatin-  
 21 ing Regional Skew.

22 If the regional and station skews differ by more  
 23 than 0.5, a careful examination of the data and the  
 24 flood-producing characteristics of the watershed should  
 25 be made. Possibly greater weight may be given to the  
 26 station skew, depending on record length, the largest  
 27 floods within the gaging record and watershed, and  
 28 watershed characteristics. Large deviations between  
 29 the regional skew and station skew may indicate that  
 30 the flood frequency characteristics of the watershed  
 31 of interest are different from those used to develop  
 32 the regional skew estimate. It is thought that station  
 33 skew is a function of rainfall skew, channel storage,  
 34 and basin storage (McCuen and Smith, 2008). There  
 35 is considerable variability of response among different  
 36 basins with similar observable characteristics, in addi-  
 37 tion to the random sampling variability in estimating  
 38 skew from a short record. It is considered reasonable  
 39 to give greater weight to the station skew, after due  
 40 consideration of the data and flood-producing charac-  
 41 teristics of the basin.

## Zeros and Identifying Potentially Influential Low Floods

Potentially influential points (“outliers”) are data  
 points ~~which-that~~ depart significantly from the trend  
 of the remaining data. In the case of annual peak  
 flows, low outliers may be floods caused by differ-  
 ent processes than the larger floods in the annual peak  
 series, as defined in the ~~Section-section~~ Zero Flows  
 and ~~Potentially-Influential~~ ~~Potentially Influential~~ Low  
 Floods. Because inclusion of these zero flow values  
 and “outliers” can significantly affect the statistical  
 parameters computed from the data, especially for  
 small samples, the presence of PILFs in the dataset  
 will bias parameter estimates.

The purpose of flood flow frequency estimation  
 is to describe the relationship between discharge and  
 exceedance probability at the high end of the fre-  
 quency distribution where AEPs are values such as  
 0.05, 0.02, 0.01, and smaller. There are cases where  
 observed values of some of the smaller annual floods  
 can have a very strong effect on the shape of the  
 estimated frequency distribution at the ~~high-discharge~~  
~~high-discharge~~ end. The purpose of the procedures  
 described here is to eliminate the influence of low  
 floods so that these small floods have little or no  
 impact on the frequency estimates at high discharges.  
 The ultimate goal is to obtain a good agreement  
 between the high end of the observed frequency dis-  
 tribution and the high end of the estimated frequency  
 distribution. This may result in a poor fit at the low  
 end of the frequency distribution ~~but-; however,~~ there  
 is generally no negative practical consequences to a  
 lack of fit at the low end.

The smallest observations in the ~~data-set-dataset~~  
 do not convey meaningful or valid information about  
 the magnitude of significant flooding (~~Appendix-appendix~~  
 6), although they do convey valid information about  
 the frequency of significant flooding. Therefore, if the  
 upper tail of the frequency curve is sensitive to the  
 numerical values of the smallest observations, then  
 that sensitivity is a spurious artifact based on the math-  
 ematical form of the assumed, ~~but in fact unknown,~~  
 flood distribution, and has no hydrologic validity.  
 Any procedure for treating outliers ultimately requires  
 judgment involving ~~both mathematical and hydrologic~~

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
 PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
 PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
 GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
 TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.



considerations. The analyst must use hydrological knowledge while applying a consistent and mathematically appropriate procedure.

These Guidelines recommend the use of a Multiple ~~Grubbs-Beck~~ Grubbs-Beck Test (MGBT) for the detection of PILFs. Statistical procedures for identifying outliers have been extensively studied, including methods for addressing the case of multiple low outliers considered here, as described in Cohn and others (2013), Lamontagne and others (2013), ~~and~~ Lamontagne and others (2016), and citations therein. The new Multiple Grubbs-Beck test was developed as an improvement to the Grubbs-Beck (GB) test (Grubbs and Beck, 1972) used in Bulletin 17B. The GB test is easily defeated by the occurrence of multiple low outliers, which exert a large distorting influence on the fitted frequency curve, but also increase the standard deviation, thereby making the standardized distances between observations too small to trigger the Grubbs-Beck test.

The MGBT is a statistically appropriate generalization of the GB test, and is sensitive to the possibility that ~~several~~ several of the smallest observations are “unusual,” or are potentially very influential. The MGBT also correctly evaluates cases where one or more observations are zero, or are below a recording threshold (partial record sites). Thus, it provides a consistent, objective, and statistically defensible algorithm that considers whether a range of the smallest observations should be classified as outliers (or PILFs) for a much wider range of situations.

The MGBT follows the same reasoning as Rosner’s R-statistic (RST) procedure (Rosner, 1983). Population mean and variance are computed from sample points that cannot be outliers under either the null or alternative hypotheses. The MGBT is a one-sided application of this procedure where only low outliers are believed to exist. In flood flow frequency analysis, high values are not treated as outliers. Low outliers are of concern because by using the logarithms of the flood peaks to fit a distribution, one or more unusual low-flow values can substantially distort the entire fitted frequency distribution. Therefore, the detection of such values is important. In addition, fitted distributions should be compared graphically with the data to check for problems.

The Multiple Grubbs-Beck test is applied to the systematic data of annual peaks from the station record. Let  $\{X_1, \dots, X_n\}$  be a series of logarithms of the annual peak floods. Consider the sorted dataset,  $\{X_{[1:n]}, X_{[2:n]}, \dots, X_{[n:n]}\}$ , where  $X_{[1:n]}$  is the smallest observation in the sample of size  $n$ . The null hypothesis ( $H_0$ ) is that all observations  $\{X_1, \dots, X_n\}$  are drawn from the same population of independent and identically distributed normal variates. The alternative hypothesis is that the  $k$ -th smallest observation in the dataset,  $X_{[k:n]}$ , is unusually small compared to that population. If  $X_{[k:n]}$  is declared a PILF, then all observations less than  $X_{[k:n]}$  are also PILFs.

Annual peaks in the ~~data set~~ dataset that are detected as potentially influential are then ~~re-coded~~ recoded as less than a threshold discharge  $T_{PILF}$  and treated as interval data in the ~~Expected Moments Algorithm~~, EMA, as discussed below. Zero flow values, if observed in the peak-flow ~~data set~~ dataset, are defined as PILFs. Computational details of the MGBT algorithm and  $p$ -values used for determining PILFs are described in ~~Appendix~~ appendix 6. For the case of a single low outlier, the Multiple Grubbs-Beck test is identical to the Grubbs-Beck test (Grubbs and Beck, 1972) that was used in IACWD (1982). Where appropriate, if the MGBT does not adequately identify PILFs, the analyst may define a low outlier threshold based on hydrological considerations, knowledge of the watershed, and site characteristics. The justification for a PILF threshold  $T_{PILF}$  should be thoroughly documented.

## Expected Moments Algorithm

The Expected Moments Algorithm (EMA) is a generalized method of moments procedure to estimate the P-III distribution parameters. The EMA provides a direct fit of the P-III distribution using the entire ~~data set~~ dataset, simultaneously employing regional skew information and a wide range of historical flood and threshold-exceedance information, while adjusting for any potentially influential low floods, missing values from an incomplete record, or zero flood years (Stedinger and Griffis, 2008). The EMA utilizes multiple types of at-site flood information, including Systematic Records, Historical Flood Information, and

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

Paleoflood and Botanical Information. It also includes information about the magnitudes of historical floods and paleofloods, flow intervals, changing base discharges from ~~crest-stage-gages~~ CSGs, and knowledge of the number of years in the historical period when no large flood occurred, as described in the ~~Section~~ section Data Representation using Flow Intervals and Perception Thresholds. The EMA also directly uses regional flood information (~~Section-section~~ Regional Information and Nearby Sites) in the form of a regional skew coefficient  $G$ . For the simple data situation that consists of a systematic record, no historical information, and no PILFs, the EMA returns the standard method of moments estimates presented in the ~~Section~~ section Parameter Estimation: Simple Case.

The EMA is the reasonable extension of the Bulletin 17B LP-III method of moments approach to deal in a consistent statistical framework with all of the sources of information that are likely to be available. There have been numerous studies that document some weaknesses and potential improvements to the moments estimation methods in Bulletin 17B, including historical data, handling of low outliers, use of regional skew, and confidence intervals. Stedinger and Cohn (1986) and Lane (1987) recognized that there are historical and paleoflood data that are not efficiently used by Bulletin 17B. The EMA was first developed as an alternative to Bulletin 17B (Lane, 1995; Lane and Cohn, 1996; Cohn and others, 1997) in order to fully use historical and paleoflood information (England and others, 2003b,a). The EMA was then extended to consistently handle low outlier adjustments and regional skew information (Griffis and others, 2004; Griffis, 2008), in addition to historical information. Confidence intervals with EMA have been developed (Cohn and others, 2001; Cohn, 2015), as described in the ~~Section-section~~ Confidence Intervals for Quantiles; thus, there is a consistent statistical framework for flood frequency. For simple cases with only a systematic record and a regional skew (see ~~Section-section~~ Parameter Estimation: Simple Case), the EMA ~~algorithm~~ reverts to the method of moments as recommended in IACWD (1982). Additional history, background, and perspectives are presented in Griffis and Stedinger (2007a) and Stedinger and Griffis (2008).

The EMA employs the peak-flow intervals  $Q_{Y,lower}$  and  $Q_{Y,upper}$  to estimate the moments of the LP-III distribution. The EMA requires the corresponding perception thresholds  $T_{Y,lower}$  and  $T_{Y,upper}$  to estimate the confidence intervals and other measures of uncertainty in frequency estimates. It is ~~therefore~~, therefore, important to estimate the flow intervals and thresholds accurately, based on all the data and information available that is presented in the ~~Section-section~~ Flood Flow Frequency Information. As described in the ~~Section-section~~ Data Representation using Flow Intervals and Perception Thresholds, peak-flow intervals and perception thresholds are defined for each data type and for each year.

For the general case of a historical perception threshold  $T_h$  and a PILF threshold  $T_{PILF}$ , the inputs to EMA are determined by counting the floods greater than ( $>$ ) (exceedances) and floods less than ( $<$ ) (censored) for each year, relative to each perception threshold. Recall that  $X = \log_{10}(Q)$ , and that  $X_h$  and  $X_{PILF}$  are the base 10 logarithms of  $T_h$  and  $T_{PILF}$ , respectively (see also the Glossary for notation and definitions). The logarithms of flood magnitudes are expressed as a union of four sets (Cohn and others, 1997), as shown below:

$$\{X\} = \{X_s^>\} \cup \{X_h^>\} \cup \{X_s^<\} \cup \{X_h^<\}, \quad (14)$$

and where PILFs are identified, the systematic period is divided into floods above and below a PILF threshold  $X_l$  (Griffis, 2008), as shown below:

$$\{X_s^<\} = \{X_l^>\} \cup \{X_l^<\} \quad (15)$$

with terms defined in ~~Table-table~~ Table 1.

The ~~Expected-Moments-Algorithm~~ EMA for the general situation with a historical flood perception threshold  $X_h$  and a PILF threshold  $X_l$  includes the following steps:

1. Perception thresholds for the historical period  $X_h$  and PILFs  $X_l$  within the systematic period are defined.
2. Using the values that exceeded the thresholds  $\{X_h^>\}$  and  $\{X_l^>\}$ , initial estimates of the sample moments  $\{\hat{\mu}, \hat{\sigma}, \hat{\tau}\}$  are computed as if one

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

**Table 1.** Flow and ~~Year~~~~year~~ terms used in ~~EMA~~~~moments~~~~the~~  
~~Expected Moments Algorithm~~.

Flow <del>Year</del> <del>year</del>	or	Definition
$\{X_s^>\}$		<del>logarithms</del> <del>=1em</del> <del>Logarithms</del> of floods that occurred in the systematic record with magnitudes that are greater than the historical threshold $X_h$ .
$\{X_h^>\}$		<del>logarithms</del> <del>=1em</del> <del>Logarithms</del> of historical floods or paleofloods with magnitudes greater than $X_h$ that occurred during the historical period.
$\{X_l^>\}$		<del>logarithms</del> <del>=1em</del> <del>Logarithms</del> of floods that occurred in the systematic record with magnitudes that are greater than the <del>PILF</del> <del>Potentially Influential Low Flood (PILF)</del> threshold $X_l$ and less than $X_h$ .
$\{X_h^<\}$		<del>logarithms</del> <del>=1em</del> <del>Logarithms</del> of unmeasured historical floods or paleofloods less than $X_h$ , because their magnitudes did not exceed $X_h$ .
$\{X_l^<\}$		<del>logarithms</del> <del>=1em</del> <del>Logarithms</del> of floods in the systematic record that are less than the PILF threshold $X_l$ .
$\{n_s^<\}$		<del>number</del> <del>=1em</del> <del>Number</del> of floods in the systematic record with magnitudes that are less than $X_h$ .
$\{n_h^<\}$		<del>number</del> <del>=1em</del> <del>Number</del> of unmeasured floods in the historical period with magnitudes that are less than $X_h$ .
$\{n_l^<\}$		<del>number</del> <del>=1em</del> <del>Number</del> of floods in the systematic record with magnitudes that are less than $X_l$ .

had a complete sample.

3. For iteration  $i = 1, 2, \dots$ , the parameters of the P-III distribution  $\{\hat{\alpha}_{i+1}, \hat{\beta}_{i+1}, \hat{\tau}_{i+1}\}$  are estimated using the previously computed sample moments, as shown in the following equations:

$$\hat{\alpha}_{i+1} = 4/\hat{\gamma}_i, \quad (16)$$

$$\hat{\beta}_{i+1} = \left(\frac{1}{2}\right) \hat{\sigma}_i \hat{\gamma}_i, \text{ and} \quad (17)$$

$$\hat{\tau}_{i+1} = \hat{\mu}_i - \hat{\alpha}_{i+1} \hat{\beta}_{i+1}. \quad (18)$$

4. New sample moments  $\{\hat{\mu}_{i+1}, \hat{\sigma}_{i+1}, \hat{\gamma}_{i+1}\}$  are estimated using expected moments.

5. Convergence test – iterate EMA steps 3 and 4 until parameter estimates converge.

For example, using the mean, as shown in equation 5, the iteration  $i + 1$  is ÷

$$\hat{\mu}_{i+1} = \left(\frac{1}{n}\right) \sum_{i=1}^n \tilde{X}_i \quad (19)$$

where

$$\tilde{X}_i = \begin{cases} X_i & \text{if } X_i \text{ is measured} \\ & \text{or “exact”} \\ E[X|X_{lower} < X_i < X_{upper}] & \text{if } X_{lower} < X_i < X_{upper} \end{cases} \quad (20)$$

and  $E[X|X_{lower} < X_i < X_{upper}]$  is the expected value of an observation known to lie within a range. The equations and computation details for EMA are presented in ~~Appendix~~~~appendix~~ 7. The EMA confidence intervals are described in the ~~Section~~~~section~~ Confidence Intervals for Quantiles.

## Record Extension with Nearby Sites

The minimum record length recommended for frequency analysis in Bulletin 17C is 10 years of annual maximum peak flows. Even with the use of an informative regional skew, historical data, and adjustment for low floods, 10 years of record may not be an adequate sample for estimating the more extreme floods like the 0.01-annual exceedance probability flood.

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

Extending records in time is a way of achieving a more representative sample. ~~There~~The following are a number of reasons why a short record station may not be representative of ~~long-term~~long-term conditions:

- the short record may represent a wet period where one or more major floods occurred in a short period of time;
- the short record may represent a drought period where no major floods occurred; and
- it may be known that large historical floods occurred prior to or after systematic data collection at the short record station and estimates of these floods need to be incorporated into the frequency analysis.

Record extension involves estimating additional years of record at the ~~short-term~~short-term station utilizing data at a nearby ~~long-term~~long-term station. The estimated annual peak flows are then analyzed along with the observed data in a Bulletin 17C frequency analysis. The recommended approach for record extension is based on the Maintenance of Variance Extension (MOVE) techniques (Hirsch, 1982) with subsequent improvements (Vogel and Stedinger, 1985). The MOVE equations, with an example application, are presented in ~~Appendix~~appendix 8. A reasonable approach to implement MOVE is to use concurrent data at a nearby ~~long-term~~long-term station that has similar watershed characteristics as the site of interest. There should be at least 10 years of overlapping data for the short record and long record stations and the correlation coefficient needs to exceed a critical value as defined in ~~Appendix~~appendix 8. It is recommended that MOVE be considered when the short record site is less than 20 years, with a minimum length of 10 years.

## Confidence Intervals for Quantiles

The user of frequency curves should be aware that the curve is only an estimate of the population curve; it is not an exact representation. A streamflow record is only a sample. How well this sample will predict the flood experience (population) depends upon the

sample size, its representativeness, and whether or not the underlying distribution is known or chosen wisely.

The record of annual peak flows at a site is a random sample of the underlying population of annual peaks and can be used to estimate the frequency curve of that population. If the same size random sample could be selected from a different period of time, a different estimate of the underlying population frequency curve probably would result. Thus, an estimated flood frequency curve can ~~be only~~only be an approximation to the true frequency curve of the underlying population of annual flood peaks. To gauge the accuracy of this approximation, one may construct an interval or range of hypothetical frequency curves that, with a high degree of confidence, contains the population frequency curve. Such intervals are called confidence intervals and their end points are called confidence limits.

Confidence intervals provide either a measure of the uncertainty of the estimated exceedance probability of a selected discharge or a measure of the uncertainty of the discharge at a selected exceedance probability. Confidence intervals on the discharge for the P-III distribution can be estimated using the method described in ~~Appendix~~appendix 7. The EMA with all available data, including historical floods, PILFs, interval data, and regional skew, is used. Uncertainty in the at-site and regional estimates of the skewness coefficients is also included.

Application of confidence intervals in reaching ~~water resource planning decision~~water resource planning decisions depends upon the needs of the user. This discussion is presented to emphasize that the frequency curve developed using this guide is only today's best estimate of the flood frequency distribution. As more data become available, the estimate will normally be improved and the confidence intervals narrowed.

## Estimating Regional Skew

As described in the ~~Section~~section Weighted Skew Coefficient Estimator, it is recommended that the skew coefficient used be a weighted average of the station skew and a regional skew (Griffis and Stedinger, 2007a). ~~A recommended procedure for esti-~~

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.



mating regional skew is using the Bayesian Weighted Least Squares/Bayesian Generalized Least Squares (B-WLS/B-GLS) method (Veilleux and others, 2011).

Tasker and Stedinger (1986) developed a weighted least squares (WLS) procedure for estimating regional skewness coefficients based on sample skewness coefficients for the logarithms of annual peak-discharge data. Their method of regional analysis of skewness estimators accounts for the precision of the skewness estimator for each station, which depends on the length of record for each station and the accuracy of an ~~Ordinary Least Squares~~ ordinary least squares (OLS) regional mean skewness. More recently, Reis and others (2005), Gruber and others (2007), and Gruber and Stedinger (2008) developed a Bayesian generalized least squares (B-GLS) regression model for regional skewness analyses. Use of a ~~GLS~~ generalized least squares (GLS) model allows the incorporation of the cross-correlation of skewness estimators. Cross-correlation arises as skewness estimators are dependent upon concurrent cross-correlation flood records. The Bayesian method allows for the computation of a posterior distribution of both the regression parameters and the model error variance. As shown in Reis and others (2005), for cases in which the model error variance is small compared to the sampling error of the at-site estimates, the Bayesian posterior distribution provides a more reasonable description of the model error variance than both the ~~generalized least squares (GLS)~~ method-of-moments GLS method of moments and maximum likelihood point estimates (Veilleux, 2011). While WLS regression accounts for the precision of the regional model and the effect of the record length on the variance of skewness coefficient estimators, GLS regression also considers the cross-correlations among the skewness coefficient estimators. The B-GLS regression procedures extend the GLS regression framework by also providing a description of the precision of the estimated model error variance, a pseudo analysis of variance, and enhanced diagnostic statistics; see also Griffis and Stedinger (2009).

Due to complexities introduced by the use of the ~~Expected Moments Algorithm (EMA)~~ EMA (Cohn and others, 1997) and large cross-correlations between annual peak discharges at some pairs of gages sites

(Parrett and others, 2011), the B-WLS/B-GLS regression procedure was developed to provide both stable and defensible results for regional skewness coefficient models (Veilleux, 2011; Veilleux and others, 2011). It uses an OLS analysis to fit an initial regional skewness model; that OLS model is then used to generate a stable regional skewness coefficient estimate for each site. That stable regional estimate is the basis for computing the variance of each at-site skewness coefficient estimator employed in the WLS analysis. Then, Bayesian WLS is used to generate estimators of the regional skewness coefficient model parameters. Finally, B-GLS is used to estimate the precision of those B-WLS parameter estimators, to estimate the model error variance and the precision of that variance estimator, and to compute various diagnostic statistics, including Bayesian plausibility values, pseudo adjusted R-squared, pseudo-Analysis of Variance table, two diagnostic error variance ratios, as well as leverage and influence metrics. This method has been successfully used to generate regional skew estimates around the ~~nation~~ Nation.

It is recommended that regional skew coefficient  $G$  estimates and mean ~~-square-square~~ error  $MSE_G$  estimates be obtained from recent studies that use the B-WLS/B-GLS regression procedure completed by the USGS. Current estimates are available for many ~~states~~ States; others are being revised by the USGS. Appendix 3 contains information regarding recent regional skew studies. Additional information on skew studies, available reports, and contact information is available at <https://acwi.gov/hydrology/Frequency/b17c/>. In lieu of current published estimates, it is recommended that users consult with the ~~U.S. Geological Survey~~ USGS to determine the availability of regional skew estimates that have been prepared using the current methods that are described in this ~~Section~~ section. The regional skew estimates published in IACWD (~~1982, Plate 1~~) IACWD (1982, plate 1) are not recommended for use in flood frequency studies.

## Comparisons of Frequency Curves

Major problems in flood frequency analysis at gaged locations are encountered when making flood

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

estimates for probabilities more rare than defined by the available record. The accuracy of flood probability estimates based upon statistical analysis of flood data deteriorates for probabilities more rare than those directly defined by the at-site flood period of record that may include systematic, historical, and paleoflood data. This is due to several major factors, including the sampling error of the statistics from the station data, because the basic underlying distribution of flood data is not exactly known, and the physical flood processes may change at larger magnitudes.

Although other procedures for estimating floods on a watershed and flood data from adjoining watersheds can sometimes be used for evaluating flood levels at high flows and rare exceedance probabilities, procedures for doing so cannot be standardized to the same extent as the procedures discussed thus far. For these situations, the Guidelines describe the information to incorporate in the analysis but allow considerable latitude in application.

Frequency curves that are estimated using the recommended procedures in the [Section section](#) Determination of the Flood Flow Frequency Curve can be compared with frequency curves from similar watersheds using regional frequency methods, or [with](#) frequency curves from precipitation using rainfall-runoff models. Independent estimates can in some cases be weighted and combined for an improved estimate as described in the [Section section](#) Weighting of Independent Frequency Estimates. Prior to making comparisons, analysts should ensure that all data and information at the location of interest and within the region, as described in the [Section section](#) Flood Flow Frequency Information and [Appendix appendix](#) 3, have been adequately considered and incorporated into the frequency analysis. In this way, the flood frequency curve may reflect (as appropriate) [the following](#): temporal information such as historical and paleoflood data; spatial information such as regional skew and watershed characteristics; and causal information such as hydroclimate information and mixed-population data. Merz and Blöschl (2008a,b) describe ways of including and combining various sources of flood frequency information.

The purpose for which the flood frequency information is needed will determine the amount of time

and effort that can justifiably be spent to obtain additional data, make comparisons with other watersheds, utilize flood estimates from precipitation, and weight the independent estimates. All types of analyses should be incorporated when estimating flood magnitudes for exceedance probabilities less than 0.01 AEP, including comparisons with similar watersheds and comparisons with flood estimates from precipitation.

The following sections describe the use of additional information to compare and potentially refine the flood frequency analysis using quantile weighting. Recommendations of specific procedures for regional comparisons or for appraising the accuracy of such estimates are beyond the scope of these Guidelines.

## Comparisons with Similar Watersheds

Comparisons and potential adjustment of a frequency curve based upon flood experience and flood statistics in nearby hydrologically similar watersheds can improve most flood frequency determinations. Use of the weighted skew coefficient recommended by these Guidelines is one form of transferring regional information to the site at hand. Additional comparisons may be helpful and are described in the following paragraphs.

A comparison between flood and extreme storm records, such as those in U.S. Army Corps of Engineers (1973) (and others) and flood flow frequency analyses at nearby hydrologically similar watersheds, will often aid in evaluating and interpreting both unusual flood experience and the flood frequency analysis of a given watershed. The shorter the flood record and the more unusual a given flood event, the greater will be the need for such comparisons.

When flood frequency curves are available for similar watersheds within a region, comparisons can be made with flood quantiles for selected exceedance probabilities or with the moments of the distribution. Flood quantile estimates from regional quantile regression models that use basin characteristics and physiographic factors, such as Paretto and others (2014b), are usually available and are recommended for use in comparisons with at-site frequency curves. Regional flood quantile methods have a long history of use (Benson, 1964, 1962; Feather and others, 2009), and

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

have been shown to perform well against alternatives (Griffis and Stedinger, 2007a). Comparisons of quantiles and frequency curve shapes can be made using the index flood method (Dalrymple, 1960; Hosking and Wallis, 1997), which may illustrate similarities or differences in flood runoff mechanisms (Bureau of Reclamation, 2002; England and others, 2010). Comparing regional moment estimates of the mean and standard deviation with at-site estimates is also informative (Griffis and Stedinger, 2007a); regional models of these moments can also be constructed (Gottwald and others, 2012). Simple drainage-area plots and peak-flow envelope curve comparisons can be useful, with an appropriate examination of flood processes and moments within a region (Blöschl and Sivapalan, 1997). If these estimates are independent of the station analysis, a weighted average of the two estimates will be more accurate than either alone. In many situations, the at-site estimate is used in a regional estimate; thus, the two estimates are correlated (Moss and Thomas, 1982).

## Comparisons with Flood Estimates from Precipitation

Floods and frequency curves developed from precipitation estimates can be used for comparison and to potentially adjust flood frequency curves, including extrapolation beyond experienced values. As described in the [Section section](#) Flood Estimates from Precipitation, flood estimates from precipitation may be available based on reconstruction of specific flood events, synthetic flood events, or continuous stream-flow estimates.

When a flood frequency curve is available from a calibrated rainfall-runoff model for the watershed of interest, comparisons can be made to estimates from the recommended procedures in the [Section section](#) Determination of the Flood Flow Frequency Curve. Plotting of the flood estimates for a range of exceedance probabilities provides a guide for potentially combining and extrapolating the frequency curve. Quantile variance estimates from the rainfall-runoff model are needed in order to potentially combine estimates. Any potential weighting or combination of frequency curves must recognize the relative accuracy of

the flood estimates and the other flood data used in the rainfall-runoff model. Whether or not such effort is warranted depends upon the procedures and data available and on the use to be made of the flood frequency estimates.

Because of the wide variety of rainfall-runoff models, parameters, and inputs, no specific procedures are recommended. Appraisal of the techniques to use flood estimates from rainfall-runoff models is currently outside the scope of these Guidelines, ~~and future work is warranted in this area. Consequently, alternative.~~ [Alternative](#) procedures for making such studies or criteria for deciding when available flood records should be combined or extended by such procedures have not been evaluated.

## Weighting of Independent Frequency Estimates

When flood frequency estimates are available from similar watersheds or from rainfall-runoff models and they are independent of the at-site estimates made using the procedures described in the [Section section](#) Determination of the Flood Flow Frequency Curve, these flood quantile estimates  $\hat{Q}_q$  may be weighted and combined. The weights are based on quantile variance and are assumed to be unbiased and independent. The weight given to each estimate is inversely proportional to its variance; ~~Appendix 9~~ [Appendix 9](#) describes the recommended weighting method and provides an example.

It is recommended that weighting be done when dependable estimates of flood quantiles and the variances of quantiles are available. Prior to weighting and combining estimates, the quantiles and variances of the estimates need to be evaluated. Flood quantile estimates may be substantially different for a variety of reasons (Rogger and others, 2012). In some situations, ~~highly-variable~~ [highly variable](#) estimates (for example, from rainfall-runoff models) may be unreliable and should not be weighted, as they would degrade the at-site estimate.

Griffis and Stedinger (2007a) evaluated several weighting methods, including quantile weighting and moment weighting with ~~2 and 3~~ [two and three](#) parameters, among ~~other alternatives.~~ [As described in](#)

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

the ~~Section~~section Regional Information and Nearby Sites, regional mean and standard deviation estimates may be available. These moments could be considered in weighting frequency curves. The computational study by Griffis and Stedinger (2007a) demonstrates that the simple weighting of at-site and regional regression quantile estimates performs nearly as well as more complex alternatives, and for short records, it provides a substantial improvement in quantile accuracy. Weighting is particularly useful when the at-site record is short (10 years). Quantile weighting, described in ~~Appendix~~appendix 9, is the recommended approach.

Analysts are encouraged to include flood frequency information from all sources, as appropriate. In some cases, information from numerous sources can be combined (Viglione and others, 2013). Other than the procedure recommended in ~~Appendix~~appendix 9, these methods have not been fully evaluated.

## Frequency Curve Extrapolation

In some situations, there is a need to estimate extreme floods with AEPs less than 0.01, such as  $Q_{0.002}$ , or other extraordinary floods. The need for these estimates may be due to an engineering design requirement, ~~floodplain analysis and management~~ (FEMA, 2015) ~~analysis, and management~~ or other infrastructure assessment. As described in the ~~Section~~section Comparisons of Frequency Curves, all types of analyses should be incorporated when estimating flood magnitudes for exceedance probabilities less than 0.01 AEP.

For these situations, the recommended approach described in the ~~Section~~section Determination of the Flood Flow Frequency Curve is appropriate, with inclusion of additional information as follows. First, expand the flood data in time for the location of interest and at sites within a region, to include historical information, paleoflood and botanical data, and extraordinary floods as described in the ~~Sections~~sections Historical Flood Information, Paleoflood and Botanical Information, and Extraordinary Floods. Additional flood data collection in the field is warranted. Second, expand and improve regional skew models using the procedures described in the ~~Section~~section

Estimating Regional Skew to include these longer records. Third, expand with regional independent information such as extreme flood rainfall-runoff models within the watershed, regional extreme flood information (frequency estimates, envelope curves, ~~etc.~~and so forth), or other physical and causal estimates as described in the ~~Section~~section Comparisons of Frequency Curves. Finally, quantify uncertainty of the quantile estimates with confidence intervals.

The amount of extrapolation depends on the quantity and quality of flood information at the site of interest, data and information within the larger region, the designs and decisions to be made, and tolerance for uncertainty in the extrapolated results. It is not simply based on the at-site data record length. ~~Because of the~~there are variations in quantity and quality of flood information, ~~and as well as in the~~ purposes of the designs and decisions to be made using the flood frequency estimates, ~~a~~a flexible approach using multiple lines of flood evidence for extrapolation is ~~recommended~~appropriate. Swain and others (2006) and Nathan and Weinmann (2016) contain additional information on extrapolation of frequency curves.

## Software and Examples

~~Specialized~~ software has been developed by various agencies that implements the recommended flood frequency procedures in these Guidelines. This includes estimating the log-Pearson Type III distribution parameters using the EMA with available historical and paleoflood data, PILFs, and regional skew information. Confidence intervals and plotting positions are also estimated. The software includes the methods and computations presented in the ~~Section~~section Determination of the Flood Flow Frequency Curve, PILFs described in ~~Appendix~~and appendix 6, and the EMA described in ~~Appendix~~appendix 7. A list of recommended software packages is provided on the HFAWG web page at <https://acwi.gov/hydrology/Frequency/b17c/>.

The initial data analysis (~~Appendix~~appendix 4) and record extension techniques (~~Appendix~~appendix 8) can be performed without the need for specialized software. Available ~~ancillary~~ materials and examples



are provided on the HFAWG web page.

Some representative flood frequency examples that illustrate the recommended methods described in the ~~Section~~ section Determination of the Flood Flow Frequency Curve are presented in ~~Appendix~~ appendix 10. The main emphasis is on the data, flow intervals, and threshold inputs to the EMA. The seven examples include the following: a systematic record; potentially ~~-influential~~ -influential low floods record; a broken record; a historical record; a crest-stage record; a historical and PILF record; and a paleoflood record. Each example includes a detailed description of the data, a ~~time-series~~ time-series plot, and a flood frequency curve. Input and output files from software used to create the examples are also available on the HFAWG web page at <https://acwi.gov/hydrology/Frequency/b17c/>. These examples are meant to illustrate the main concepts presented in these Guidelines, and are not meant to be ~~all-inclusive~~ all inclusive.

## Future Studies

These Guidelines are designed to meet a current, ever-pressing demand that the Federal Government develop a coherent set of procedures for accurately defining flood potentials as needed in programs of flood damage abatement. Much additional study and data are required before the twin goals of accuracy and consistency will be obtained. It is hoped that this guide contributes to this effort by defining the essential elements of a coherent set of procedures for flood frequency determination. Although selection of the analytical procedures to be used in each step or element of the analysis has been carefully made based upon a review of the literature, the considerable practical experience of Work Group members, and special studies conducted to aid in the selection process, ~~the~~ there is a need for additional studies ~~is recognized~~.

The following is a list of some additional needed topics of study identified by the Work Group:

1. ~~the~~ The identification and treatment of mixed distributions, including those based on hydrometeorological or hydrological conditions;

2. ~~guides~~ Guides for defining flood potentials for ungaged watersheds and watersheds with limited gaging records, as described below;
3. ~~methods~~ Methods to include watershed hydrological processes and physical considerations into the analysis that can influence the frequency curve;
4. ~~procedures~~ Procedures for improving flood frequency analysis using precipitation data, rainfall-runoff models, and associated uncertainty analysis;
5. ~~guides~~ Guides for defining flood potentials for watersheds altered by urbanization, wildfires, deforestation, and by reservoirs, as described below;
6. ~~guides~~ Guides for estimating dynamic flood frequency curves that vary with time, incorporating climate indices, changing basin characteristics, and addressing potential nonstationary climate conditions;
7. ~~frequency~~ Frequency estimation in cases where long-term trends are evident in the data but are not readily explainable by the history of land use, land use practices, or engineering modifications of the river or ~~floodplain~~ flood plain; and
8. ~~an~~ An examination and redefinition of risk, reliability, and return periods under nonstationary conditions.

There is a need to develop guidance in the following three important areas: ungaged sites, regulated flow frequency, and urbanization. Some existing practices are listed below for each area; ~~however no specific recommendations and guidance is made.~~ While significant work has been done on these topics by researchers around the world, those efforts have not yet been evaluated for broad and systematic application as contemplated in these Guidelines.

## Ungaged Sites

Many of the stream sites of interest do not have gages with ~~sufficient records~~ provisional or are ungaged. One

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

area of future work needed is to develop national guidance on methods for estimating flood flow frequency curves at ungaged sites. ~~Two common methods that~~ The following two common methods are used to estimate frequency curves for ungaged watersheds ~~are~~ (Thomas and others, 2001): (1) regional flood quantile regression equations based on generalized least squares (Tasker and Stedinger, 1989); and (2) rainfall-runoff models (Pilgrim and Cordery, 1993; McCuen, 2004). Regional regression equations are available through the USGS StreamStats software (Ries and others, 2008). A limited comparison of these two methods is in Thomas and others (2001).

## Regulated Flow Frequency

A large portion of the stream sites of interest have flows that are altered to some degree by regulating structures such as dams, reservoirs, and diversions, or flows are affected by levees. One area of future work needed is to develop national guidance on methods for estimating flood flow frequency curves at stream locations affected by varying degrees of regulation. Some common regulated flood frequency methods include estimating unregulated flows using empirical relationships or synthetic floods (U.S. Army Corps of Engineers, 1993), graphical frequency analysis, or by applying total probability concepts (Kubik, 1990; Sanders and others, 1990). Durrans (2002) summarizes these approaches and describes other methods that could be considered.

## Urbanization and Watershed Change

At many stream sites of interest, flood frequency relationships may be changing due to alterations within watershed and the stream corridor over time. This may be due to urbanization (Konrad, 2003), land development, and other factors described in the ~~Section~~ section Watershed Changes. National guidance for estimating flood flow frequency curves in watersheds experiencing urbanization and ~~for~~(or) watershed change is an area needing further work. One option is to develop flood frequency regression equations that include urbanization factors (Sauer and others, 1983). Other approaches for estimating flood frequency for water-

sheds undergoing land use change are in McCuen (2003) and Villarini and others (2009b).

## Applicability of These Guidelines

Bulletin 17C goes a long way towards addressing known concerns with Bulletin 17B. However, many concerns remain, such as the best methods of addressing regulated flows and mixed distributions, methods for addressing urbanizing areas and other ~~land-use~~ land use changes, better ways to use information provided by rainfall records and rainfall-frequency analyses, and better use of physiographic watershed characteristics to define the flood flow frequency relationship. How to handle climate change and climate variability will ~~be continuing concerns~~ continue to be issues of concern as science comes to better understand the likely impact of such atmospheric phenomena on hydrologic processes. Development of flood flow frequency relationships between gaged and ungaged sites is an important topic not addressed here.

While many improvements have been made, there are significant limitations that apply to use of procedures ~~recommended~~ included in these Guidelines. First and foremost, these Guidelines are predicated on the availability of flood data that constitute a reliable, representative, and homogeneous sample of expected future floods. Flood data that represent unique occurrences such as dam failures, ice jams, or importation or diversion of flood waters should not be used to characterize flood potential unless they are properly adjusted to represent prevailing (natural) watershed conditions. There are currently many concerns about potential changes in the distribution of floods due to watershed changes and anthropogenic climate change; such concerns may require special procedures as discussed in the ~~Section~~ section Data Assumptions and Specific Concerns.

These Guidelines assume the use of the annual-maximum flood series and generally apply only to portions of the flood frequency curve for AEPs less than 0.10. Flood frequencies for larger, more common AEPs may be more appropriately determined from use of the partial-duration series ~~data, that~~ (PDS) data, which allow for ~~more than one large flood per year~~

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.



rather than the annual-maximum flood series. Some procedures for these analyses are mentioned in the ~~Section~~section Flood Flow Frequency Information.

These Guidelines apply only to those situations for which there are sufficient data for carrying out the necessary computations. In general, flood frequency computations are not reliable with records comprised of less than 10 annual flood observations. Accurate determination of floods for small AEPs ( $<0.01$ ) generally requires more data; estimations of floods for AEPs smaller than 0.005 generally require augmentation of the systematically observed flood records with general regional information, insight from precipitation records, or paleoflood information, as available (~~Section~~section Flood Flow Frequency Information).

These Guidelines permit augmentation of flood records by incorporation of community experience such as the documentation of floods in news reports, community accounts, or paleoflood indicators (see the ~~Sections~~sections Historical Flood Information and Paleoflood and Botanical Information and ~~Appendix~~appendix 3). However, these conditions must be properly described by specification of accurate observation intervals and thresholds based upon consideration of the physical flood indicators and hydraulic conditions. These considerations must be well documented by a qualified analyst, together with the necessary computations.

The Guidelines may be used to estimate flood frequencies for urban conditions where there are flood observation datasets of sufficient length that represent stable development or that can be adjusted to account for changes in urban infrastructure and routing parameters (Sauer and others, 1983; McCuen, 2003). Similarly, any regional skewness estimator should be derived from flood records representing urban conditions.

These Guidelines describe the set of procedures recommended for defining flood potential as expressed by a flood flow frequency curve. Special situations may require other approaches, perhaps defining the frequency relationship for flood volumes or river stages. In those cases where the procedures of this guide are not followed, deviations must be supported by appropriate study, including a comparison of the results obtained with those obtained using the ~~recommended~~

included procedures.

There is much concern about changes in flood risk associated with climate variability and long-term climate change. Time invariance was assumed in the development of this guide. In those situations where there is sufficient scientific evidence to facilitate quantification of the impact of climate variability or change in flood risk, this knowledge should be incorporated in flood frequency analysis by employing time-varying parameters or other appropriate techniques. All such methods employed need to be thoroughly documented and justified.

It is not anticipated that many special situations warranting other approaches will occur at sites ~~which~~that have reasonable flood flow records. These procedures should be followed, unless there are compelling technical reasons for departing from the Guidelines. These deviations are to be documented and supported by appropriate study, including the comparison of results. The Subcommittee on Hydrology requests that these situations be called to its attention for consideration in future modifications of these Guidelines.

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

## References

- Abramowitz, M., and Stegun, I.A., 1964, Handbook of Mathematical Functions: Washington, National Bureau of Standards.
- Aldridge, B.N., and Eychaner, J.H., 1984, Floods of October 1977 in southern Arizona and March 1978 in central Arizona: Reston, VA, U.S. Geological Survey Water-Supply Paper 2243, 143 p.
- Aldridge, B.N., and Hales, T.A., 1984, Floods of November 1978 to March 1979 in Arizona and West-Central New Mexico: Reston, VA, U.S. Geological Survey Water-Supply Paper 2241, 149 p.
- Alila, Younes, and Mtiraoui, Ahmed, 2002, Implications of heterogeneous flood-frequency distributions on traditional stream-discharge prediction techniques: *Hydrological Processes*, v. 16, no. 5, p. 1065–1084, accessed August 29, 2017, at <https://doi.org/10.1002/hyp.346>.
- Asquith, W.H., and Slade, R.M., 1995, Documented and potential extreme peak discharges and relation between potential extreme peak discharges and probable maximum flood peak discharges in Texas, U.S. Geological Survey, Reston, VA, 58 p.
- Baker, V.R., 2013, Global Late Quaternary Fluvial Paleohydrology: With Special Emphasis on Paleofloods and Megafloods, *in* Shroder, J.F., and Wohl, E., v. 9.26, *Treatise on Geomorphology*: San Diego, Academic Press, 511 - 527 p., accessed August 29, 2017, at <https://doi.org/10.1016/B978-0-12-374739-6.00252-9>.
- Baker, V.R., 1987, Paleoflood hydrology and extraordinary flood events: *Journal of Hydrology*, v. 96, no. 1-4, p. 79–99.
- Baker, V.R., 2008, Paleoflood hydrology: Origin, progress, prospects: *Geomorphology*, v. 101, no. 1-2, p. 1–13.
- Baker, V.R., Kochel, R.C., and Patton, P.C., 1988, *Flood Geomorphology*: New York, John Wiley and Sons, 503 p.
- Baker, V.R., Webb, R.H., and House, P.K., 2002, The scientific and societal value of paleoflood hydrology, *in* House, P.K., Webb, R.H., Baker, V.R., and Levish, D.R., *Water Science and Application Series*, v. 5, *Ancient Floods, Modern Hazards*, American Geophysical Union, 1-19 p., accessed August 29, 2017, at <https://doi.org/10.1029/WS005p0001>.
- Bathurst, J., 1986, Slope-area discharge gaging in mountain rivers: *Journal of Hydraulic Engineering*, v. 112, no. 5, p. 376–391, accessed August 27, 2017, at [https://doi.org/10.1061/\(ASCE\)0733-9429\(1986\)112:5\(376\)](https://doi.org/10.1061/(ASCE)0733-9429(1986)112:5(376)).
- Beard, L.R., 1974, *Flood flow frequency techniques*, Center for Research in Water Resources.
- Benito, G., and O'Connor, J.E., 2013, Quantitative Paleoflood Hydrology, *in* Shroder, J.F., and Wohl, E., v. 9.24, *Treatise on Geomorphology*: San Diego, Academic Press, 459 - 474 p., accessed June 27, 2017, at <https://doi.org/10.1016/B978-0-12-374739-6.00250-5>.
- Benson, M.A., 1962, Factors Influencing the Occurrence of Floods in a Humid Region of Diverse Terrain, U.S. Geological Survey, Reston, VA, 64 p.
- Benson, M.A., 1964, Factors Affecting the Occurrence of Floods in the Southwest, U.S. Geological Survey, Reston, VA, 72 p.
- Benson, M.A., 1968, Uniform flood-frequency estimating methods for federal agencies: *Water Resources Research*, v. 4, no. 5, p. 891–908, accessed August 29, 2017, at <https://doi.org/10.1029/WR004i005p00891>.
- Benson, M.A., and Dalrymple, Tate, 1967, General field and office procedures for indirect discharge measurements, U.S. Geological Survey, 30 p.
- Beven, K.J., 2001, *Rainfall-Runoff Modeling*, The Primer: Chichester, John Wiley and Son, 360 p.
- Blainey, J, Webb, R, Moss, M, and Baker, V, 2002, Bias and information content of paleoflood data in flood-frequency analysis: *Water Science and Application*, v. 5, p. 161–174.
- Blöschl, Günter, and Sivapalan, Murugesu, 1997, Process controls on regional flood frequency: Coefficient of variation and basin scale: *Water Resources Research*, v. 33, no. 12, p. 2967–2980,

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

- accessed August 29, 2017, at <https://doi.org/10.1029/97WR00568>.
- Brekke, L.D., Kiang, J.E., Olsen, J.R., Pulwarty, R.S., Raff, D.A., Turnipseed, D.P., Webb, R.S., and White, K.D., 2009, Climate change and water resources management – A Federal perspective: Reston, VA, U.S. Geological Survey Circular 1331, 65 p.
- Bureau of Reclamation, 2002, Flood Hazard Analysis - Folsom Dam, Central Valley Project, California, U.S. Department of Interior, Bureau of Reclamation, Denver, CO, 128 p.
- Clark, Martyn, Gangopadhyay, Subhrendu, Hay, Lauren, Rajagopalan, Balaji, and Wilby, Robert, 2004, The Schaake shuffle: a method for reconstructing space-time variability in forecasted precipitation and temperature fields: *Journal of Hydrometeorology*, v. 5, no. 1, p. 243–262, accessed August 29, 2017, at [https://doi.org/10.1175/1525-7541\(2004\)005\(0243:TSSAMF\)2.0.CO;2](https://doi.org/10.1175/1525-7541(2004)005(0243:TSSAMF)2.0.CO;2).
- Cohen, A.C., 1991, Truncated and Censored Samples - Theory and Application, Marcel-Dekker, New York, 312 p.
- Cohn, T.A., 2015, A fast numerical algorithm for computing accurate confidence intervals for EMA-based flood quantile estimates: (to be submitted to *Water Resources Research*).
- Cohn, T.A., and Lins, H.F., 2005, Nature's style: Naturally trendy: *Geophysical Research Letters*, v. 32, no. 23, p. n/a–n/a, accessed August 30, 2017, at <https://doi.org/10.1029/2005GL024476>.
- Cohn, T.A., Lane, W.L., and Baier, W.G., 1997, An algorithm for computing moments-based flood quantile estimates when historical flood information is available: *Water Resources Research*, v. 33, no. 9, p. 2089–2096.
- Cohn, T.A., Lane, W.L., and Stedinger, J.R., 2001, Confidence intervals for Expected Moments Algorithm flood quantile estimates: *Water Resources Research*, v. 37, no. 6, p. 1695–1706.
- Cohn, T.A., England, J.F., Berenbrock, C.E., Mason, R.R., Stedinger, J.R., and Lamontagne, J.R., 2013, A generalized grubbs-beck test statistic for detecting multiple potentially influential low outliers in flood series: *Water Resources Research*, v. 49, no. 8, p. 5047–5058, accessed June 27, 2017, at <https://doi.org/10.1002/wrcr.20392>.
- Cohn, T.A., Barth, N.A., England, J.F., Faber, B., Mason, R.R., and Stedinger, J.R., 2017, Evaluation of Recommended Revision to Bulletin 17B, U.S. Geological Survey, Reston, VA, 183 p.
- Coles, Stuart, 2001, An Introduction to Statistical Modeling of Extreme Values, Springer, 208 p., accessed August 28, 2017, at <https://doi.org/10.1007/978-1-4471-3675-0>.
- Condie, R., and Lee, K.A., 1982, Flood frequency analysis with historic information: *Journal of Hydrology*, v. 58, p. 47–61.
- Cook, J.L., 1987, Quantifying peak discharges for historical floods: *Journal of Hydrology*, v. 96, no. 14, p. 29 – 40, accessed August 30, 2017, at [https://doi.org/10.1016/0022-1694\(87\)90141-7](https://doi.org/10.1016/0022-1694(87)90141-7).
- Costa, J.E., 1978, Holocene stratigraphy in flood frequency analysis: *Water Resources Research*, v. 14, no. 4, p. 626–632, accessed August 30, 2017, at <https://doi.org/10.1029/WR014i004p00626>.
- Costa, J.E., 1986, A history of paleoflood hydrology in the United States, 1800-1970: *EOS, Transactions of the American Geophysical Union*, v. 67, no. 17, p. 425, 428–430.
- Costa, J.E., 1987, A history of paleoflood hydrology in the United States, 1800-1970, in Gillmor, C.S., Landa, E.R., Ince, Simon, and Back, William, v. 3, History of Geophysics: Volume 3, American Geophysical Union, 49–53 p., accessed August 30, 2017, at <https://doi.org/10.1029/HG003p0049>.
- Costa, J.E., and Baker, V.R., 1981, *Surficial Geology: Building With the Earth*, Wiley, 498 p.
- Costa, J.E., and Jarrett, R.D., 2008, An evaluation of selected extraordinary floods in the United States reported by the U.S. Geological Survey and implications for future advancement of flood science, U.S. Geological Survey, Reston, VA, 232 p.
- Crippen, J.R., 1978, Composite Log-Type III Frequency-Magnitude Curve of Annual Floods, U.S. Geological Survey, 5 p.

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

- Crippen, J.R., and Bue, C.D., 1977, Maximum flood flows in the conterminous United States: Reston, VA, U.S. Geological Survey Water-Supply Paper 1887, 52 p.
- Crowfoot, R.M., Paillett, A.V., Ritz, G.F., Smith, M.E., Jenkins, R.A., and O'Neill, G.B., 1997, Water resources data, Colorado, Water Year 1996, volume 2. Colorado River Basin., U.S. Geological Survey, Reston, VA, 551 p.
- Dalrymple, Tate, 1960, Flood-frequency analyses. manual of hydrology: Part 3. flood flow techniques, U.S. Geological Survey, Reston, VA, 80 p.
- Dalrymple, Tate, and Benson, M.A., 1967, Measurement of peak discharge by the slope-area method, U.S. Geological Survey, 12 p.
- David, H.A., 1981, Order Statistics, Wiley, 360 p.
- Duan, Qingyun, Gupta, H.V., Sorooshian, Soroosh, Rousseau, A.N., and Turcotte, Richard, 2003, Calibration of Watershed Models, American Geophysical Union, 345 p., accessed August 30, 2017, at <https://doi.org/10.1029/WS006>.
- Durlin, R.R., and Schaffstall, W.P., 2002, Water resources data, Pennsylvania, Water Year 2002, volume 2. Susquehanna River Basin., U.S. Geological Survey, Reston, VA, 536 p.
- Durrans, S.R., 2002, Regulated flood frequency methods, informal handouts, Hydrologic Frequency Analysis Work Group, Hydrology Subcommittee, Interagency Advisory Committee on Water Data.
- Eash, D.A., 2010, Estimating flood frequency, in Mutel, C.F., A Watershed Year: Anatomy of the Iowa Floods of 2008: Iowa City, University of Iowa Press, 61-70 p.
- Eash, D.A., Barnes, K.K., and Veilleux, A.G., 2013, Methods for Estimating Annual Exceedance-Probability Discharges for Streams in Iowa, Based on Data through Water Year 2010, U.S. Geological Survey, Reston, VA, 63 p.
- Elliott, J.G., Jarrett, R.D., and Ebling, J.L., 1982, Annual snowmelt and rainfall peak-flow data on selected foothills region streams, South Platte River, Arkansas River, and Colorado River Basins, Colorado, U.S. Geological Survey, 86 p.
- England, J.F., 1998, Assessment of historical and paleohydrologic information in flood frequency analysis, M.S. thesis, Colorado State University, 292 p.
- England, J.F., and Cohn, T.A., 2007, Scientific and practical considerations related to revising Bulletin 17B: The case for improved treatment of historical information and low outliers, in American Society of Civil Engineers, EWRI World Water & Environmental Resources Congress.
- England, J.F., and Cohn, T.A., 2008, Bulletin 17B Flood Frequency Revisions: Practical Software and Test Comparison Results, in World Environmental and Water Resources Congress 2008 Ahupua, ASCE, p. 1–11.
- England, J.F., Jarrett, R.D., and Salas, J.D., 2003a, Data-based comparisons of moments estimators using historical and paleoflood data: Journal of Hydrology, v. 278, no. 1, p. 172–196.
- England, J.F., Salas, J.D., and Jarrett, R.D., 2003b, Comparisons of two moments-based estimators that utilize historical and paleoflood data for the log Pearson type III distribution.: Water Resources Research, v. 39, no. 9, p. 1243.
- England, J.F., Klawon, J.E., Klinger, R.E., and Bauer, T.R., 2006, Flood hazard analysis for Pueblo dam, Colorado, final report, U.S. Department of Interior, Bureau of Reclamation, Denver, CO, 160 p.
- England, J.F., Godaire, J.E., Klinger, R.E., Bauer, T.R., and Julien, P.Y., 2010, Paleohydrologic bounds and extreme flood frequency of the Upper Arkansas River, Colorado, USA: Geomorphology, v. 124, no. 1, p. 1–16.
- Feaster, T.D., Gotvald, A.J., and Weaver, J.C., 2009, Magnitude and Frequency of Rural Floods in the Southeastern United States, 2006: Volume 3, South Carolina, U.S. Geological Survey, Reston, VA, 226 p.
- FEMA, 2009, Guidelines and Specifications for Flood Hazard Mapping Partners. Appendix C: Guidance for Riverine Flooding Analyses and

DRAFT: August 31, 2017

PROVISIONAL  
THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.



- Mapping, Federal Emergency Management Agency, 77 p.
- FEMA, 2015, Guidelines for Implementing Executive Order 11988, Floodplain Management, and Executive Order 13690, Establishing a Federal Flood Risk Management Standard and a Process for Further Soliciting and Considering Stakeholder Input, Federal Emergency Management Agency, 80 p.
- Follansbee, R., and Jones, E.E., 1922, The Arkansas River floods of June 3-5, 1921, U.S. Geological Survey, Reston, VA, 44 p.
- Follansbee, R., and Sawyer, L.R., 1948, Floods in Colorado, U.S. Geological Survey, Reston, VA, 151 p.
- Gerard, R., and Karpuk, E.W., 1979, Probability analysis of historical flood data: *Journal of Hydraulic Engineering*, v. 105, no. HY9, p. 1153–1165.
- Godaire, J.E., and Bauer, T.R., 2012, Paleoflood Study, North Fork Red River Basin near Altus Dam, Oklahoma, U.S. Department of Interior, Bureau of Reclamation, Denver, CO, 55 p.
- Godaire, J.E., and Bauer, T.R., 2013, Paleoflood Study on the Rio Chama near El Vado Dam, New Mexico, U.S. Department of Interior, Bureau of Reclamation, Denver, CO, 58 p.
- Godaire, J.E., Bauer, T.R., and Klinger, R.E., 2012, Paleoflood Study, San Joaquin River near Friant Dam, California, U.S. Department of Interior, Bureau of Reclamation, Denver, CO, 60 p.
- Gotvald, A.J., Feaster, T.D., and Weaver, J.C., 2009, Magnitude and Frequency of Rural Floods in the Southeastern United States, 2006: Volume 1, Georgia, U.S. Geological Survey, Reston, VA, 120 p.
- Gotvald, A.J., Barth, N.A., Veilleux, A.G., and Parrett, Charles, 2012, Methods for Determining Magnitude and Frequency of Floods in California, Based on Data through Water Year 2006, U.S. Geological Survey, Reston, VA, 38 p.
- Griffis, V.W., and Stedinger, J.R., 2007a, Evolution of Flood Frequency Analysis with Bulletin 17: *Journal of Hydrologic Engineering*, v. 12, no. 3, p. 283–297.
- Griffis, V.W., and Stedinger, J.R., 2007b, Log-Pearson Type 3 Distribution and Its Application in Flood Frequency Analysis. I: Distribution Characteristics: *Journal of Hydrologic Engineering*, v. 12, no. 5, p. 482–491.
- Griffis, V.W., and Stedinger, J.R., 2007c, Log-Pearson Type 3 Distribution and Its Application in Flood Frequency Analysis. II: Parameter Estimation Methods: *Journal of Hydrologic Engineering*, v. 12, no. 5, p. 492–500, accessed August 30, 2017, at [https://doi.org/10.1061/\(ASCE\)1084-0699\(2007\)12:5\(492\)](https://doi.org/10.1061/(ASCE)1084-0699(2007)12:5(492)).
- Griffis, V., and Stedinger, J., 2009, Log-Pearson Type 3 Distribution and Its Application in Flood Frequency Analysis. III: Sample Skew and Weighted Skew Estimators: *Journal of Hydrologic Engineering*, v. 14, no. 2, p. 121–130, accessed August 30, 2017, at [https://doi.org/10.1061/\(ASCE\)1084-0699\(2009\)14:2\(121\)](https://doi.org/10.1061/(ASCE)1084-0699(2009)14:2(121)).
- Griffis, V.W., 2008, EMA with Historical Information, Low Outliers, and Regional Skew, *in* World Environmental and Water Resources Congress 2008 Ahupua, ASCE, p. 1–10.
- Griffis, V.W., and Stedinger, J.R., 2007d, The use of GLS regression in regional hydrologic analyses: *Journal of Hydrology*, v. 344, no. 1-2, p. 82–95.
- Griffis, V.W., Stedinger, J.R., and Cohn, T.A., 2004, Log Pearson type 3 quantile estimators with regional skew information and low outlier adjustments.: *Water Resources Research*, v. 40, no. 7, p. W07,503.
- Grover, N.C., 1937, The floods of March 1936, Part 3, Potomac, James, and upper Ohio Rivers: Reston, VA, U.S. Geological Survey Water-Supply Paper 800, 351 p.
- Grover, N.C., and Harrington, A.W., 1943, Stream Flow: New York, John Wiley and Sons, 363 p.
- Grubbs, F.E., and Beck, G., 1972, Extension of sample sizes and percentage points for significance tests of outlying observations: *Technometrics*, v. 14, no. 4, p. 847–854.
- Gruber, A.M., and Stedinger, J.R., 2008, Models of LP3 regional skew, data selection, and Bayesian



- GLS regression, *in* World Environmental & Water Resources Conference 2008 AHUPUA'A.
- Gruber, A.M., Reis, D.S., and Stedinger, J.R., 2007, Models of regional skew based on Bayesian GLS regression, *in* World Environmental & Water Resources Conference, Tampa, FL, edited by K. C. Kabbes, ASCE EWRI.
- Harden, T.M., O'Connor, J.E., Driscoll, D.G., and Stamm, J.F., 2011, Flood-frequency analyses from paleoflood investigations for Spring, Rapid, Boxelder, and Elk Creeks, Black Hills, western South Dakota, U.S. Geological Survey, Reston, VA, 136 p.
- Hazen, Allen, 1930, Flood Flows: New York, John Wiley and Sons, 199 p.
- Helsel, D.R., and Hirsch, R.M., 1992, Statistical methods in water resources, Elsevier, 529 p.
- Hirsch, R.M., 1982, A comparison of four stream-flow record extension techniques: Water Resources Research, v. 18, no. 4, p. 1081–1088.
- Hirsch, R.M., 1987, Probability plotting position formulas for flood records with historical information: Journal of Hydrology, v. 96, no. 1-4, p. 185–199.
- Hirsch, R.M., and DeCicco, L.A., 2015, User Guide to Exploration and Graphics for RivEr Trends (EGRET) and dataRetrieval: R Packages for Hydrologic Data, U.S. Geological Survey, 93 p.
- Hirsch, R.M., and Stedinger, J.R., 1987, Plotting positions for historical floods and their precision: Water Resources Research, v. 23, no. 4, p. 715–727.
- Hirsch, R.M., Helsel, D.R., Cohn, T.A., and Gilroy, E.J., 1993, Statistical analysis of hydrologic data, *in* Maidment, D.R., Handbook of Hydrology, McGraw Hill, Inc., 17.1-17.55 p.
- Hirschboeck, K.K., 1987a, Catastrophic flooding and atmospheric circulation anomalies, *in* Mayer, L., and Nash, D., Catastrophic Flooding: Boston, Allen and Unwin, 23-56 p.
- Hirschboeck, K.K., 1987b, Hydroclimatically-defined mixed distributions in partial duration flood series, *in* Hydrologic Frequency Modeling,
- edited by V. P. Singh, Louisiana State University, Baton Rouge, D. Reidel Publishing Company, p. 199–212.
- Hirschboeck, K.K., 1991, Climate and floods, *in* Paulson, R.W., Chase, E.B., Roberts, R.S., and Moody, D.W., National Water Summary 1988-89-Hydrologic Events and Floods and Droughts, U.S. Geological Survey Water-Supply Paper 2375: Reston, VA, U.S. Geological Survey, 67-88 p.
- Hosking, J. R.M., and Wallis, J.R., 1997, Regional Frequency Analysis – An Approach Based on L-Moments, Cambridge University Press, 224 p.
- House, P.K., and Baker, V.R., 2001, Paleohydrology of flash floods in small desert watersheds in western Arizona: Water Resources Research, v. 37, no. 6, p. 1825–1839.
- House, P.K., Webb, R.H., Baker, V.R., and Levish, D.R., 2002a, Ancient Floods, Modern Hazards, American Geophysical Union, 385 p., accessed August 30, 2017, at <https://doi.org/10.1029/WS005>.
- House, P.K., Pearthree, P.A., and Klawon, J.E., 2002b, Historical flood and paleoflood chronology of the lower Verde River, Arizona: Stratigraphic evidence and related uncertainties, *in* House, P.K., Webb, R.H., Baker, V.R., and Levish, D.R., Water Science and Application Series, v. 5, Ancient Floods, Modern Hazards, American Geophysical Union, 267–293 p., accessed August 30, 2017, at <https://doi.org/10.1029/WS005p0267>.
- Hupp, C.R., 1987, Botanical Evidence of Floods and Paleoflood History, *in* Regional Flood Frequency Analysis, edited by V. P. Singh, D. Reidel, p. 355–369.
- Hupp, C.R., 1988, Plant ecological aspects of flood geomorphology and paleoflood history, *in* Baker, V.R., Kochel, R.C., and Patton, P.C., Flood Geomorphology, John Wiley and Sons, 335-356 p.
- Hupp, C.R., and Osterkamp, W.R., 1996, Riparian vegetation and fluvial geomorphic processes: Geomorphology, v. 14, no. 4, p. 277–295.
- Hydrologic Frequency Analysis Work Group, 2006, Flood Frequency Research Needs. Memorandum

DRAFT: August 31, 2017

PROVISIONAL  
THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

- to Subcommittee on Hydrology, accessed August 30, 2017, at [http://acwi.gov/hydrology/Frequency/minutes/HFAWG\\_for\\_revising\\_Bulletin17B\\_4.pdf](http://acwi.gov/hydrology/Frequency/minutes/HFAWG_for_revising_Bulletin17B_4.pdf).
- Hydrologic Frequency Analysis Work Group, 2013, Recommendations Memorandum to the Subcommittee on Hydrology, Hydrologic Frequency Analysis Work Group Meeting Minutes, accessed August 30, 2017, at [http://acwi.gov/hydrology/Frequency/minutes/HFAWG\\_mins\\_june12-13\\_2013\\_meeting\\_revised.071213.pdf](http://acwi.gov/hydrology/Frequency/minutes/HFAWG_mins_june12-13_2013_meeting_revised.071213.pdf).
- IACWD, 1982, Guidelines for determining flood flow frequency, Bulletin 17-B, Interagency Committee on Water Data, Hydrology Subcommittee.
- Jain, Shaleen, and Lall, Upmanu, 2001, Floods in a changing climate: Does the past represent the future?: *Water Resources Research*, v. 37, no. 12, p. 3193–3205, accessed August 30, 2017, at <https://doi.org/10.1029/2001WR000495>.
- Jarrett, R.D., 1987, Errors in slope-area computations of peak discharges in mountain streams: *Journal of Hydrology*, v. 96, no. 14, p. 53 – 67, accessed August 30, 2017, at [https://doi.org/10.1016/0022-1694\(87\)90143-0](https://doi.org/10.1016/0022-1694(87)90143-0).
- Jarrett, R.D., 1991, Paleohydrology and its value in analyzing floods and droughts, in Paulson, R.W., Chase, E.B., Roberts, R.S., and Moody, D.W., National Water Summary 1988-89-Hydrologic Events and Floods and Droughts, U.S. Geological Survey Water-Supply Paper 2375: Reston, VA, U.S. Geological Survey, 105-116 p.
- Jarrett, R.D., and Costa, J.E., 1988, Evaluation of the flood hydrology in the Colorado Front Range using precipitation, streamflow, and paleoflood data for the Big Thompson River Basin, U.S. Geological Survey, Reston, VA, 37 p.
- Jarrett, R.D., and England, J.F., 2002, Reliability of paleostage indicators for paleoflood studies, in House, P.K., Webb, R.H., Baker, V.R., and Levish, D.R., Water Science and Application Series, v. 5, Ancient Floods, Modern Hazards, American Geophysical Union, 91-109 p., accessed August 30, 2017, at <https://doi.org/10.1029/WS005p0091>.
- Jarrett, R.D., and Malde, H.E., 1987, Paleodischarge of the late pleistocene Bonneville flood, Snake River, Idaho, computed from new evidence: *Geological Society of America Bulletin*, v. 99, p. 127–134.
- Jarrett, R.D., and Tomlinson, E.M., 2000, Regional interdisciplinary paleoflood approach to assess extreme flood potential: *Water Resources Research*, v. 36, no. 10, p. 2957–2984.
- Kiang, J.E., Olsen, J.R., and Waskom, R.M., 2011, Introduction to the featured collection on “non-stationarity, hydrologic frequency analysis, and water management”: *JAWRA Journal of the American Water Resources Association*, v. 47, no. 3, p. 433–435, accessed August 30, 2017, at <https://doi.org/10.1111/j.1752-1688.2011.00551.x>.
- Kirby, W., 1972, Computer-oriented Wilson-Hilferty transformation that preserves the first three moments and the lower bound of the Pearson type 3 distribution: *Water Resources Research*, v. 8, no. 5, p. 1251–1254, accessed August 30, 2017, at <https://doi.org/10.1029/WR008i005p01251>.
- Kirby, W.H., 1987, Linear error analysis of slope-area discharge determinations: *Journal of Hydrology*, v. 96, no. 14, p. 125 – 138, accessed August 30, 2017, at [https://doi.org/https://dx.doi.org/10.1016/0022-1694\(87\)90148-X](https://doi.org/https://dx.doi.org/10.1016/0022-1694(87)90148-X).
- Kite, G.W., 1988, Frequency and Risk Analyses in Hydrology: Littleton, Colorado, Water Resources Publications, 258 p.
- Kite, J.S., Gebhardt, T.W., and Springer, G.S., 2002, Slackwater deposits as paleostage indicators in canyon reaches of the Central Appalachians: Reevaluation after the 1996 Cheat River flood, in House, P.K., Webb, R.H., Baker, V.R., and Levish, D.R., Water Science and Application Series, v. 5, Ancient Floods, Modern Hazards, American Geophysical Union, 257–266 p., accessed August 30, 2017, at <https://doi.org/10.1029/WS005p0257>.
- Kjeldsen, T.R., Lamb, Rob, and Blazkova, S.D., 2014, Uncertainty in flood frequency analysis, in Beven, Keith, and Hall, Jim, Applied Uncertainty Analysis for Flood Risk Management, Imperial College Press, 153-197 p., accessed August 30, 2017, at <https://doi.org/10.1017/9781107300000.008>.

DRAFT: August 31, 2017

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

#### 44 Guidelines for Determining Flood Flow Frequency – Bulletin 17C

- August 30, 2017, at <https://doi.org/10.1142/9781848162716.0008>.
- Klemeš, V., 1986, Dilettantism in hydrology: Transition or destiny?: *Water Resources Research*, v. 22, no. 9, p. 177S–188S.
- Klemeš, V., 1987, Hydrological and engineering relevance of flood frequency analysis, *in* *Hydrologic Frequency Modeling*, edited by V. P. Singh, Louisiana State University, Baton Rouge, D. Reidel Publishing Company, p. 1–18.
- Klemeš, V., 2000, Tall Tales about Tails of Hydrological Distributions. I: *Journal of Hydrologic Engineering*, v. 5, no. 3, p. 227–231, accessed August 30, 2017, at [https://doi.org/10.1061/\(ASCE\)1084-0699\(2000\)5:3\(227\)](https://doi.org/10.1061/(ASCE)1084-0699(2000)5:3(227)).
- Klinger, R.E., and Bauer, T.R., 2010, Paleoflood Study on the South Fork of the Boise River for Anderson Ranch Dam, Idaho, U.S. Department of Interior, Bureau of Reclamation, Denver, CO, 30 p.
- Klinger, R.E., and Godaire, J.E., 2002, Development of a Paleoflood Database for Rivers in the Western U.S., U.S. Department of Interior, Bureau of Reclamation, Denver, CO, 38 p.
- Kochel, R.C., Baker, V.R., and Patton, P.C., 1982, Paleohydrology of southwestern Texas: *Water Resources Research*, v. 18, no. 4, p. 1165–1183.
- Koenig, T.A., Bruce, J.L., O'Connor, J.E., McGee, B.D., Holmes, R.R., Hollins, Ryan, Forbes, B.T., Kohn, M.S., Schellekens, M.F., Martin, Z.W., and Peppler, M.C., 2016, Identifying and preserving high-water mark data, U.S. Geological Survey, 47 p.
- Kohn, M.S., Jarrett, R.D., Krammes, G.S., and Mommandi, Amanullah, 2013, Web-based flood database for Colorado, water years 1867 through 2011, U.S. Geological Survey, 26 p.
- Konrad, C.P., 2003, Effects of urban development on floods, U.S. Geological Survey, 4 p.
- Koutsoyiannis, Demetris, 2011, Hurst-Kolmogorov dynamics and uncertainty: *JAWRA Journal of the American Water Resources Association*, v. 47, no. 3, p. 481–495, accessed August 30, 2017, at <https://doi.org/10.1111/j.1752-1688.2011.00543.x>.
- Kubik, H.E., 1990, Computation of Regulated Frequency Curves by Application of the Total Probability Theorem, U.S. Army Corps of Engineers, Hydrologic Engineering Center, 17 p.
- Kuczera, G., 1982, Robust flood frequency models: *Water Resources Research*, v. 18, no. 2, p. 315–324, accessed August 30, 2017, at <https://doi.org/10.1029/WR018i002p00315>.
- Kuczera, G., 1996, Correlated rating curve error in flood frequency inference: *Water Resources Research*, v. 32, no. 7, p. 2119–2127.
- Kuichling, E., 1917, Discussion of flood flows by W.E. Fuller: *Transactions, ASCE*, v. 77, no. 1293, p. 643–663.
- Lamontagne, J.R., Stedinger, J.R., Berenbrock, Charles, Veilleux, A.G., Ferris, J.C., and Knifong, D.L., 2012, Development of Regional Skews for Selected Flood Durations for the Central Valley Region, California, Based on Data Through Water Year 2008, U.S. Geological Survey, Reston, VA, 60 p.
- Lamontagne, J.R., Stedinger, J.R., Cohn, T.A., and Barth, N.A., 2013, Robust national flood frequency guidelines: What is an outlier?, *in* *Showcasing the Future*, ASCE, p. 2454–2466, accessed August 30, 2017, at <https://doi.org/10.1061/9780784412947.242>.
- Lamontagne, J.R., Stedinger, J.R., Yu, Xin, Whealton, C.A., and Xu, Ziyao, 2016, Robust flood frequency analysis: Performance of EMA with multiple Grubbs-Beck outlier tests: *Water Resources Research*, v. 52, no. 4, p. 3068–3084, accessed August 30, 2017, at <https://doi.org/10.1002/2015WR018093>.
- Lane, W.L., 1987, Paleohydrologic data and flood frequency estimation, *in* *Application of Frequency and Risk in Water Resources*, edited by V. P. Singh, D. Reidel, p. 287–298.
- Lane, W.L., 1995, Method of moments approach to historical data, informal handout, Bulletin 17B

DRAFT: August 31, 2017

PROVISIONAL  
THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.  
IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).  
IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

- Working Group, Hydrology Subcommittee, Interagency Advisory Committee on Water Data.
- Lane, W.L., and Cohn, T.A., 1996, Expected moments algorithms for flood frequency analysis, *in* North American Water and Environment Congress & Destructive Water, ASCE, p. 2185–2190.
- Lang, M., Ouarda, T.B.M.J., and Bobée, B., 1999, Towards operational guidelines for over-threshold modeling: *Journal of Hydrology*, v. 225, no. 3–4, p. 103–117, accessed August 30, 2017, at [https://doi.org/10.1016/S0022-1694\(99\)00167-5](https://doi.org/10.1016/S0022-1694(99)00167-5).
- Langbein, W.B., 1949, Annual floods and the partial-duration flood series: *Transactions of the American Geophysical Union*, v. 30, no. 6, p. 879–881.
- Langbein, W.B., and Iseri, K.T., 1960, General introduction and hydrologic definitions. manual of hydrology: Part 1. general surface water techniques, U.S. Geological Survey, Reston, VA, 29 p.
- Leese, M.N., 1973, Use of censored data in the estimation of gumbel distribution parameters for annual maximum flood series: *Water Resources Research*, v. 9, no. 6, p. 1534–1542.
- Levish, D.R., 2002, Non-exceedance information for flood hazard assessment, *in* House, P.K., Webb, R.H., Baker, V.R., and Levish, D.R., *Water Science and Application Series*, v. 5, *Ancient Floods, Modern Hazards*, American Geophysical Union, 91–109 p., accessed August 30, 2017, at <https://doi.org/10.1029/WS005p0175>.
- Levish, D.R., England, J.F., Klawon, J.E., and O’Connell, D. R.H., 2003, Flood hazard analysis for Seminole and Glendo dams, Kendrick and North Platte projects, Wyoming, final report, U.S. Department of Interior, Bureau of Reclamation, Denver, CO, 126 p.
- Lichty, R.W., and Liscum, Fred, 1978, A rainfall-runoff modeling procedure for improving estimates of T-year (annual) floods for small drainage basins, U.S. Geological Survey, Reston, VA, 44 p.
- Lins, H.F., and Cohn, T.A., 2011, Stationarity: Wanted dead or alive?: *JAWRA Journal of the American Water Resources Association*, v. 47, no. 3, p. 475–480, accessed August 30, 2017, at <https://doi.org/10.1111/j.1752-1688.2011.00542.x>.
- Lombard, F., 1987, Rank tests for changepoint problems: *Biometrika*, v. 74, no. 3, p. 615–624, accessed August 30, 2017, at <https://doi.org/10.1093/biomet/74.3.615>.
- Maddox, R.A., Canova, F., and Hoxit, L.R., 1980, Meteorological characteristics of flash flood events over the Western United States: *Monthly Weather Review*, v. 108, no. 11, p. 1866–1877.
- Madsen, Henrik, Rasmussen, P.F., and Rosbjerg, Dan, 1997, Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 1. At-site modeling: *Water Resources Research*, v. 33, no. 4, p. 747–757, accessed August 30, 2017, at <https://doi.org/10.1029/96WR03848>.
- Mallakpour, Iman, and Villarini, Gabriele, 2016, A simulation study to examine the sensitivity of the Pettitt test to detect abrupt changes in mean: *Hydrological Sciences Journal*, v. 61, no. 2, p. 245–254, accessed August 30, 2017, at <https://doi.org/10.1080/02626667.2015.1008482>.
- Mastin, M.C., 2007, Re-evaluation of the 1921 peak discharge at Skagit River near Concrete, Washington, U.S. Geological Survey, Reston, VA, 12 p.
- Matalas, N.C., and Jacobs, B., 1964, A Correlation Procedure for Augmenting Hydrologic Data, U.S. Geological Survey, 7 p.
- Matthai, H.F., 1969, Floods of June 1965 in South Platte River Basin, Colorado: Washington, DC, U.S. Geological Survey Water-Supply Paper 1850-B, 64 p.
- May, J.E., Gorman, J.G., Goodrich, R.D., Bobier, M.W., and Miller, V.E., 1996, Water resources data, Iowa, Water Year 1995, U.S. Geological Survey, Reston, VA, 387 p.
- McCabe, G.J., and Wolock, D.M., 2002, A step increase in streamflow in the conterminous United States: *Geophysical Research Letters*, v. 29, no. 24, p. 38–1–38–4, accessed August 30, 2017, at <https://doi.org/10.1029/2002GL015999>.
- McCord, V. A. S., 1990, Augmenting flood frequency

DRAFT: August 31, 2017

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.



- estimates using flood-scarred trees, Ph.D. dissertation, University of Arizona, 182 p. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2009\)14:4\(303\)](https://doi.org/10.1061/(ASCE)1084-0699(2009)14:4(303)).
- McCuen, R., and Knight, Z., 2006, Fuzzy analysis of slope-area discharge estimates: *Journal of Irrigation and Drainage Engineering*, v. 132, no. 1, p. 64–69, accessed August 30, 2017, at [https://doi.org/10.1061/\(ASCE\)0733-9437\(2006\)132:1\(64\)](https://doi.org/10.1061/(ASCE)0733-9437(2006)132:1(64)).
- McCuen, R.H., 2003, *Modeling Hydrologic Change: Statistical Methods*: Boca Raton, Lewis Publishers, 433 p.
- McCuen, R.H., 2004, *Hydrologic Analysis and Design*: Upper Saddle River, NJ, Prentice Hall, 888 p.
- McCuen, R.H., and Smith, Eric, 2008, Origin of flood skew: *Journal of Hydrologic Engineering*, v. 13, no. 9, p. 771–775, accessed August 30, 2017, at [https://doi.org/10.1061/\(ASCE\)1084-0699\(2008\)13:9\(771\)](https://doi.org/10.1061/(ASCE)1084-0699(2008)13:9(771)).
- McGlashan, H.D., and Briggs, R.C., 1939, *Floods of December 1937 in Northern California*, U.S. Geological Survey, Reston, VA, 497 p.
- Merz, R., and Blöschl, G., 2003, A process typology of regional floods: *Water Resources Research*, v. 39, no. 12, p. SWC 5–1 – SWC 5–20, accessed August 30, 2017, at <https://doi.org/10.1029/2002WR001952>.
- Merz, R., and Blöschl, G., 2008a, Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information: *Water Resources Research*, v. 44, no. 8, accessed August 30, 2017, at <https://doi.org/10.1029/2007WR006744>.
- Merz, R., and Blöschl, G., 2008b, Flood frequency hydrology: 2. Combining data evidence: *Water Resources Research*, v. 44, no. 8, accessed August 30, 2017, at <https://doi.org/10.1029/2007WR006745>.
- MGS Engineering Consultants, 2009, *General Storm Stochastic Event Flood Model (SEFM) - Technical Support Manual*, Bureau of Reclamation, Denver, CO.
- Moglen, G.E., 2009, Hydrology and impervious areas: *Journal of Hydrologic Engineering*, v. 14, no. 4, p. 303–304, accessed August 30, 2017, at [https://doi.org/10.1061/\(ASCE\)1084-0699\(2009\)14:4\(303\)](https://doi.org/10.1061/(ASCE)1084-0699(2009)14:4(303)).
- Moglen, G.E., and Beighley, R.E., 2002, Spatially explicit hydrologic modeling of land use change: *JAWRA Journal of the American Water Resources Association*, v. 38, no. 1, p. 241–252, accessed August 30, 2017, at <https://doi.org/10.1111/j.1752-1688.2002.tb01548.x>.
- Moglen, G.E., and Shivers, D.E., 2006, Methods for adjusting U.S. Geological Survey rural regression peak discharges in an urban setting, U.S. Geological Survey, Reston, VA, 55 p.
- Moran, M.A., 1974, On estimators obtained from a sample augmented by multiple regression: *Water Resources Research*, v. 10, no. 1, p. 81–85.
- Moss, M.E., and Thomas, W.O., 1982, Discussion of “WRC standard flood frequency guidelines” by D.R. Jackson: *J. Water Resources Planning and Management*, v. 108, no. 1, p. 166–168.
- Munn, J., and Savage, J.L., 1922, *The Flood of June 1921 in the Arkansas River*, at Pueblo, Colorado: *Transactions, ASCE*, v. 85, no. 1480, p. 1–65.
- Murphy, P., 2001, Evaluation of mixed-population flood-frequency analysis: *Journal of Hydrologic Engineering*, v. 6, no. 1, p. 62–70, accessed August 30, 2017, at [https://doi.org/10.1061/\(ASCE\)1084-0699\(2001\)6:1\(62\)](https://doi.org/10.1061/(ASCE)1084-0699(2001)6:1(62)).
- Nathan, R.J., and Weinmann, E., 2016, *Estimation of Very Rare to Extreme Floods*, Engineers Australia, accessed August 30, 2017, at <http://arr.ga.gov.au/>.
- National Research Council, 1988, *Estimating probabilities of extreme floods*, National Research Council, 141 p.
- National Research Council, 1995, *Flood risk management and the American River basin: an evaluation*, National Academy Press.
- National Research Council, 1999, *Improving American River flood frequency analyses*, National Research Council, 123 p.
- National Research Council, 2000, *Risk analysis and uncertainty in flood damage reduction studies*, National Academy Press, 202 p.

DRAFT: August 31, 2017

PROVISIONAL  
THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.



- Nayak, M.A., and Villarini, Gabriele, 2016, Evaluation of the capability of the Lombard test in detecting abrupt changes in variance: *Journal of Hydrology*, v. 534, p. 451 – 465, accessed August 30, 2017, at <https://doi.org/10.1016/j.jhydrol.2016.01.016>.
- O'Connor, J.E., and Costa, J.E., 2004, Spatial distribution of the largest rainfall-runoff floods from basins between 2.6 and 26,000 km<sup>2</sup> in the United States and Puerto Rico: *Water Resources Research*, v. 40, no. 1, accessed August 30, 2017, at <https://doi.org/10.1029/2003WR002247>, w01107.
- O'Connor, J.E., Webb, R.H., and Baker, V.R., 1986, Paleohydrology of pool-and-riffle pattern development: Boulder Creek, Utah: *Geological Society of America Bulletin*, v. 97, no. 4, p. 410–420, accessed August 30, 2017, at [https://doi.org/10.1130/0016-7606\(1986\)97<410:POPPDB>2.0.CO;2](https://doi.org/10.1130/0016-7606(1986)97<410:POPPDB>2.0.CO;2).
- O'Connor, J.E., Atwater, B.F., Cohn, T.A., Cronin, T.M., Keith, M.K., Smith, C.G., and Mason, R.R., 2014, Assessing Inundation Hazards to Nuclear Powerplant Sites Using Geologically Extended Histories of Riverine Floods, Tsunamis, and Storm Surges, U.S. Geological Survey, Reston, VA, 65 p.
- Olsen, J.R., Kiang, Julie, and Waskom, Reagan, 2010, Workshop on Nonstationarity, Hydrologic Frequency Analysis, and Water Management, Colorado Water Institute, Colorado State University, Fort Collins, CO, 304 p.
- Olson, S.A., 2014, Estimation of flood discharges at selected annual exceedance probabilities for unregulated, rural streams in Vermont, with a section on Vermont regional skew regression, by Veilleux, A.G., U.S. Geological Survey, Reston, VA, 27 p.
- Osterkamp, W.R., and Costa, J.E., 1987, Changes accompanying an extraordinary flood on a sand-bed stream, *in* Mayer, L., and Nash, D., *Catastrophic Flooding*: Boston, Allen and Unwin, 201–224 p.
- Over, T.M., and Soong, D.T., 2015, Flood-Peak Magnitudes at Selected Stations in Northeastern Illinois Adjusted for Changes in Land-Use Conditions, U.S. Geological Survey, Reston, VA.
- Parretti, N.V., Kennedy, J.R., and Cohn, T.A., 2014a, Evaluation of the expected moments algorithm and a multiple low-outlier test for flood frequency analysis at streamgaging stations in Arizona, U.S. Geological Survey, Reston, VA, 61 p.
- Parretti, N.V., Kennedy, J.R., Turney, L.A., and Veilleux, A.G., 2014b, Methods for estimating magnitude and frequency of floods in Arizona, developed with unregulated and rural peak-flow data through water year 2010, U.S. Geological Survey, Reston, VA, 61 p.
- Parrett, Charles, Veilleux, Andrea, Stedinger, J.R., Barth, N.A., Knifong, D.L., and Ferris, J.C., 2011, Regional skew for California, and flood frequency for selected sites in the Sacramento–San Joaquin River Basin, based on data through Water Year 2006, U.S. Geological Survey, Reston, VA, 94 p.
- Pettitt, A.N., 1979, A non-parametric approach to the change-point problem: *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, v. 28, no. 2, p. 126–135, accessed August 30, 2017, at <http://www.jstor.org/stable/2346729>.
- Pilgrim, D.H., and Cordery, I., 1993, Flood runoff, *in* Maidment, D.R., *Handbook of Hydrology*, McGraw Hill, Inc., 9.1–9.72 p.
- Potter, K.W., and Walker, J.F., 1985, An empirical study of flood measurement error: *Water Resources Research*, v. 21, no. 3, p. 403–406.
- Prior, J.C., 1991, *Landforms of Iowa*, University of Iowa Press, 154 p.
- Quessy, J.-F., Favre, A.-C., Said, Meriem, and Champagne, Maryse, 2011, Statistical inference in Lombard's smooth-change model: *Environmetrics*, v. 22, no. 7, p. 882–893, accessed August 30, 2017, at <https://doi.org/10.1002/env.1108>.
- Quick, M.C., 1991, Reliability of flood discharge estimates: *Canadian Journal of Civil Engineering*, v. 18, p. 624–630.
- Rantz, S.E., and Others, 1982a, Measurement and computation of streamflow: Volume 1, Measurement of stage and discharge: Reston, VA, U.S. Geological Survey, 1982a, *Measurement and computation of streamflow: Volume 1, Measurement of stage and discharge*: Reston, VA, U.S. Geological Survey, 1982a, 1982b, 1982c, 1982d, 1982e, 1982f, 1982g, 1982h, 1982i, 1982j, 1982k, 1982l, 1982m, 1982n, 1982o, 1982p, 1982q, 1982r, 1982s, 1982t, 1982u, 1982v, 1982w, 1982x, 1982y, 1982z, 1983a, 1983b, 1983c, 1983d, 1983e, 1983f, 1983g, 1983h, 1983i, 1983j, 1983k, 1983l, 1983m, 1983n, 1983o, 1983p, 1983q, 1983r, 1983s, 1983t, 1983u, 1983v, 1983w, 1983x, 1983y, 1983z, 1984a, 1984b, 1984c, 1984d, 1984e, 1984f, 1984g, 1984h, 1984i, 1984j, 1984k, 1984l, 1984m, 1984n, 1984o, 1984p, 1984q, 1984r, 1984s, 1984t, 1984u, 1984v, 1984w, 1984x, 1984y, 1984z, 1985a, 1985b, 1985c, 1985d, 1985e, 1985f, 1985g, 1985h, 1985i, 1985j, 1985k, 1985l, 1985m, 1985n, 1985o, 1985p, 1985q, 1985r, 1985s, 1985t, 1985u, 1985v, 1985w, 1985x, 1985y, 1985z, 1986a, 1986b, 1986c, 1986d, 1986e, 1986f, 1986g, 1986h, 1986i, 1986j, 1986k, 1986l, 1986m, 1986n, 1986o, 1986p, 1986q, 1986r, 1986s, 1986t, 1986u, 1986v, 1986w, 1986x, 1986y, 1986z, 1987a, 1987b, 1987c, 1987d, 1987e, 1987f, 1987g, 1987h, 1987i, 1987j, 1987k, 1987l, 1987m, 1987n, 1987o, 1987p, 1987q, 1987r, 1987s, 1987t, 1987u, 1987v, 1987w, 1987x, 1987y, 1987z, 1988a, 1988b, 1988c, 1988d, 1988e, 1988f, 1988g, 1988h, 1988i, 1988j, 1988k, 1988l, 1988m, 1988n, 1988o, 1988p, 1988q, 1988r, 1988s, 1988t, 1988u, 1988v, 1988w, 1988x, 1988y, 1988z, 1989a, 1989b, 1989c, 1989d, 1989e, 1989f, 1989g, 1989h, 1989i, 1989j, 1989k, 1989l, 1989m, 1989n, 1989o, 1989p, 1989q, 1989r, 1989s, 1989t, 1989u, 1989v, 1989w, 1989x, 1989y, 1989z, 1990a, 1990b, 1990c, 1990d, 1990e, 1990f, 1990g, 1990h, 1990i, 1990j, 1990k, 1990l, 1990m, 1990n, 1990o, 1990p, 1990q, 1990r, 1990s, 1990t, 1990u, 1990v, 1990w, 1990x, 1990y, 1990z, 1991a, 1991b, 1991c, 1991d, 1991e, 1991f, 1991g, 1991h, 1991i, 1991j, 1991k, 1991l, 1991m, 1991n, 1991o, 1991p, 1991q, 1991r, 1991s, 1991t, 1991u, 1991v, 1991w, 1991x, 1991y, 1991z, 1992a, 1992b, 1992c, 1992d, 1992e, 1992f, 1992g, 1992h, 1992i, 1992j, 1992k, 1992l, 1992m, 1992n, 1992o, 1992p, 1992q, 1992r, 1992s, 1992t, 1992u, 1992v, 1992w, 1992x, 1992y, 1992z, 1993a, 1993b, 1993c, 1993d, 1993e, 1993f, 1993g, 1993h, 1993i, 1993j, 1993k, 1993l, 1993m, 1993n, 1993o, 1993p, 1993q, 1993r, 1993s, 1993t, 1993u, 1993v, 1993w, 1993x, 1993y, 1993z, 1994a, 1994b, 1994c, 1994d, 1994e, 1994f, 1994g, 1994h, 1994i, 1994j, 1994k, 1994l, 1994m, 1994n, 1994o, 1994p, 1994q, 1994r, 1994s, 1994t, 1994u, 1994v, 1994w, 1994x, 1994y, 1994z, 1995a, 1995b, 1995c, 1995d, 1995e, 1995f, 1995g, 1995h, 1995i, 1995j, 1995k, 1995l, 1995m, 1995n, 1995o, 1995p, 1995q, 1995r, 1995s, 1995t, 1995u, 1995v, 1995w, 1995x, 1995y, 1995z, 1996a, 1996b, 1996c, 1996d, 1996e, 1996f, 1996g, 1996h, 1996i, 1996j, 1996k, 1996l, 1996m, 1996n, 1996o, 1996p, 1996q, 1996r, 1996s, 1996t, 1996u, 1996v, 1996w, 1996x, 1996y, 1996z, 1997a, 1997b, 1997c, 1997d, 1997e, 1997f, 1997g, 1997h, 1997i, 1997j, 1997k, 1997l, 1997m, 1997n, 1997o, 1997p, 1997q, 1997r, 1997s, 1997t, 1997u, 1997v, 1997w, 1997x, 1997y, 1997z, 1998a, 1998b, 1998c, 1998d, 1998e, 1998f, 1998g, 1998h, 1998i, 1998j, 1998k, 1998l, 1998m, 1998n, 1998o, 1998p, 1998q, 1998r, 1998s, 1998t, 1998u, 1998v, 1998w, 1998x, 1998y, 1998z, 1999a, 1999b, 1999c, 1999d, 1999e, 1999f, 1999g, 1999h, 1999i, 1999j, 1999k, 1999l, 1999m, 1999n, 1999o, 1999p, 1999q, 1999r, 1999s, 1999t, 1999u, 1999v, 1999w, 1999x, 1999y, 1999z, 2000a, 2000b, 2000c, 2000d, 2000e, 2000f, 2000g, 2000h, 2000i, 2000j, 2000k, 2000l, 2000m, 2000n, 2000o, 2000p, 2000q, 2000r, 2000s, 2000t, 2000u, 2000v, 2000w, 2000x, 2000y, 2000z, 2001a, 2001b, 2001c, 2001d, 2001e, 2001f, 2001g, 2001h, 2001i, 2001j, 2001k, 2001l, 2001m, 2001n, 2001o, 2001p, 2001q, 2001r, 2001s, 2001t, 2001u, 2001v, 2001w, 2001x, 2001y, 2001z, 2002a, 2002b, 2002c, 2002d, 2002e, 2002f, 2002g, 2002h, 2002i, 2002j, 2002k, 2002l, 2002m, 2002n, 2002o, 2002p, 2002q, 2002r, 2002s, 2002t, 2002u, 2002v, 2002w, 2002x, 2002y, 2002z, 2003a, 2003b, 2003c, 2003d, 2003e, 2003f, 2003g, 2003h, 2003i, 2003j, 2003k, 2003l, 2003m, 2003n, 2003o, 2003p, 2003q, 2003r, 2003s, 2003t, 2003u, 2003v, 2003w, 2003x, 2003y, 2003z, 2004a, 2004b, 2004c, 2004d, 2004e, 2004f, 2004g, 2004h, 2004i, 2004j, 2004k, 2004l, 2004m, 2004n, 2004o, 2004p, 2004q, 2004r, 2004s, 2004t, 2004u, 2004v, 2004w, 2004x, 2004y, 2004z, 2005a, 2005b, 2005c, 2005d, 2005e, 2005f, 2005g, 2005h, 2005i, 2005j, 2005k, 2005l, 2005m, 2005n, 2005o, 2005p, 2005q, 2005r, 2005s, 2005t, 2005u, 2005v, 2005w, 2005x, 2005y, 2005z, 2006a, 2006b, 2006c, 2006d, 2006e, 2006f, 2006g, 2006h, 2006i, 2006j, 2006k, 2006l, 2006m, 2006n, 2006o, 2006p, 2006q, 2006r, 2006s, 2006t, 2006u, 2006v, 2006w, 2006x, 2006y, 2006z, 2007a, 2007b, 2007c, 2007d, 2007e, 2007f, 2007g, 2007h, 2007i, 2007j, 2007k, 2007l, 2007m, 2007n, 2007o, 2007p, 2007q, 2007r, 2007s, 2007t, 2007u, 2007v, 2007w, 2007x, 2007y, 2007z, 2008a, 2008b, 2008c, 2008d, 2008e, 2008f, 2008g, 2008h, 2008i, 2008j, 2008k, 2008l, 2008m, 2008n, 2008o, 2008p, 2008q, 2008r, 2008s, 2008t, 2008u, 2008v, 2008w, 2008x, 2008y, 2008z, 2009a, 2009b, 2009c, 2009d, 2009e, 2009f, 2009g, 2009h, 2009i, 2009j, 2009k, 2009l, 2009m, 2009n, 2009o, 2009p, 2009q, 2009r, 2009s, 2009t, 2009u, 2009v, 2009w, 2009x, 2009y, 2009z, 2010a, 2010b, 2010c, 2010d, 2010e, 2010f, 2010g, 2010h, 2010i, 2010j, 2010k, 2010l, 2010m, 2010n, 2010o, 2010p, 2010q, 2010r, 2010s, 2010t, 2010u, 2010v, 2010w, 2010x, 2010y, 2010z, 2011a, 2011b, 2011c, 2011d, 2011e, 2011f, 2011g, 2011h, 2011i, 2011j, 2011k, 2011l, 2011m, 2011n, 2011o, 2011p, 2011q, 2011r, 2011s, 2011t, 2011u, 2011v, 2011w, 2011x, 2011y, 2011z, 2012a, 2012b, 2012c, 2012d, 2012e, 2012f, 2012g, 2012h, 2012i, 2012j, 2012k, 2012l, 2012m, 2012n, 2012o, 2012p, 2012q, 2012r, 2012s, 2012t, 2012u, 2012v, 2012w, 2012x, 2012y, 2012z, 2013a, 2013b, 2013c, 2013d, 2013e, 2013f, 2013g, 2013h, 2013i, 2013j, 2013k, 2013l, 2013m, 2013n, 2013o, 2013p, 2013q, 2013r, 2013s, 2013t, 2013u, 2013v, 2013w, 2013x, 2013y, 2013z, 2014a, 2014b, 2014c, 2014d, 2014e, 2014f, 2014g, 2014h, 2014i, 2014j, 2014k, 2014l, 2014m, 2014n, 2014o, 2014p, 2014q, 2014r, 2014s, 2014t, 2014u, 2014v, 2014w, 2014x, 2014y, 2014z, 2015a, 2015b, 2015c, 2015d, 2015e, 2015f, 2015g, 2015h, 2015i, 2015j, 2015k, 2015l, 2015m, 2015n, 2015o, 2015p, 2015q, 2015r, 2015s, 2015t, 2015u, 2015v, 2015w, 2015x, 2015y, 2015z, 2016a, 2016b, 2016c, 2016d, 2016e, 2016f, 2016g, 2016h, 2016i, 2016j, 2016k, 2016l, 2016m, 2016n, 2016o, 2016p, 2016q, 2016r, 2016s, 2016t, 2016u, 2016v, 2016w, 2016x, 2016y, 2016z, 2017a, 2017b, 2017c, 2017d, 2017e, 2017f, 2017g, 2017h, 2017i, 2017j, 2017k, 2017l, 2017m, 2017n, 2017o, 2017p, 2017q, 2017r, 2017s, 2017t, 2017u, 2017v, 2017w, 2017x, 2017y, 2017z, 2018a, 2018b, 2018c, 2018d, 2018e, 2018f, 2018g, 2018h, 2018i, 2018j, 2018k, 2018l, 2018m, 2018n, 2018o, 2018p, 2018q, 2018r, 2018s, 2018t, 2018u, 2018v, 2018w, 2018x, 2018y, 2018z, 2019a, 2019b, 2019c, 2019d, 2019e, 2019f, 2019g, 2019h, 2019i, 2019j, 2019k, 2019l, 2019m, 2019n, 2019o, 2019p, 2019q, 2019r, 2019s, 2019t, 2019u, 2019v, 2019w, 2019x, 2019y, 2019z, 2020a, 2020b, 2020c, 2020d, 2020e, 2020f, 2020g, 2020h, 2020i, 2020j, 2020k, 2020l, 2020m, 2020n, 2020o, 2020p, 2020q, 2020r, 2020s, 2020t, 2020u, 2020v, 2020w, 2020x, 2020y, 2020z, 2021a, 2021b, 2021c, 2021d, 2021e, 2021f, 2021g, 2021h, 2021i, 2021j, 2021k, 2021l, 2021m, 2021n, 2021o, 2021p, 2021q, 2021r, 2021s, 2021t, 2021u, 2021v, 2021w, 2021x, 2021y, 2021z, 2022a, 2022b, 2022c, 2022d, 2022e, 2022f, 2022g, 2022h, 2022i, 2022j, 2022k, 2022l, 2022m, 2022n, 2022o, 2022p, 2022q, 2022r, 2022s, 2022t, 2022u, 2022v, 2022w, 2022x, 2022y, 2022z, 2023a, 2023b, 2023c, 2023d, 2023e, 2023f, 2023g, 2023h, 2023i, 2023j, 2023k, 2023l, 2023m, 2023n, 2023o, 2023p, 2023q, 2023r, 2023s, 2023t, 2023u, 2023v, 2023w, 2023x, 2023y, 2023z, 2024a, 2024b, 2024c, 2024d, 2024e, 2024f, 2024g, 2024h, 2024i, 2024j, 2024k, 2024l, 2024m, 2024n, 2024o, 2024p, 2024q, 2024r, 2024s, 2024t, 2024u, 2024v, 2024w, 2024x, 2024y, 2024z, 2025a, 2025b, 2025c, 2025d, 2025e, 2025f, 2025g, 2025h, 2025i, 2025j, 2025k, 2025l, 2025m, 2025n, 2025o, 2025p, 2025q, 2025r, 2025s, 2025t, 2025u, 2025v, 2025w, 2025x, 2025y, 2025z, 2026a, 2026b, 2026c, 2026d, 2026e, 2026f, 2026g, 2026h, 2026i, 2026j, 2026k, 2026l, 2026m, 2026n, 2026o, 2026p, 2026q, 2026r, 2026s, 2026t, 2026u, 2026v, 2026w, 2026x, 2026y, 2026z, 2027a, 2027b, 2027c, 2027d, 2027e, 2027f, 2027g, 2027h, 2027i, 2027j, 2027k, 2027l, 2027m, 2027n, 2027o, 2027p, 2027q, 2027r, 2027s, 2027t, 2027u, 2027v, 2027w, 2027x, 2027y, 2027z, 2028a, 2028b, 2028c, 2028d, 2028e, 2028f, 2028g, 2028h, 2028i, 2028j, 2028k, 2028l, 2028m, 2028n, 2028o, 2028p, 2028q, 2028r, 2028s, 2028t, 2028u, 2028v, 2028w, 2028x, 2028y, 2028z, 2029a, 2029b, 2029c, 2029d, 2029e, 2029f, 2029g, 2029h, 2029i, 2029j, 2029k, 2029l, 2029m, 2029n, 2029o, 2029p, 2029q, 2029r, 2029s, 2029t, 2029u, 2029v, 2029w, 2029x, 2029y, 2029z, 2030a, 2030b, 2030c, 2030d, 2030e, 2030f, 2030g, 2030h, 2030i, 2030j, 2030k, 2030l, 2030m, 2030n, 2030o, 2030p, 2030q, 2030r, 2030s, 2030t, 2030u, 2030v, 2030w, 2030x, 2030y, 2030z, 2031a, 2031b, 2031c, 2031d, 2031e, 2031f, 2031g, 2031h, 2031i, 2031j, 2031k, 2031l, 2031m, 2031n, 2031o, 2031p, 2031q, 2031r, 2031s, 2031t, 2031u, 2031v, 2031w, 2031x, 2031y, 2031z, 2032a, 2032b, 2032c, 2032d, 2032e, 2032f, 2032g, 2032h, 2032i, 2032j, 2032k, 2032l, 2032m, 2032n, 2032o, 2032p, 2032q, 2032r, 2032s, 2032t, 2032u, 2032v, 2032w, 2032x, 2032y, 2032z, 2033a, 2033b, 2033c, 2033d, 2033e, 2033f, 2033g, 2033h, 2033i, 2033j, 2033k, 2033l, 2033m, 2033n, 2033o, 2033p, 2033q, 2033r, 2033s, 2033t, 2033u, 2033v, 2033w, 2033x, 2033y, 2033z, 2034a, 2034b, 2034c, 2034d, 2034e, 2034f, 2034g, 2034h, 2034i, 2034j, 2034k, 2034l, 2034m, 2034n, 2034o, 2034p, 2034q, 2034r, 2034s, 2034t, 2034u, 2034v, 2034w, 2034x, 2034y, 2034z, 2035a, 2035b, 2035c, 2035d, 2035e, 2035f, 2035g, 2035h, 2035i, 2035j, 2035k, 2035l, 2035m, 2035n, 2035o, 2035p, 2035q, 2035r, 2035s, 2035t, 2035u, 2035v, 2035w, 2035x, 2035y, 2035z, 2036a, 2036b, 2036c, 2036d, 2036e, 2036f, 2036g, 2036h, 2036i, 2036j, 2036k, 2036l, 2036m, 2036n, 2036o, 2036p, 2036q, 2036r, 2036s, 2036t, 2036u, 2036v, 2036w, 2036x, 2036y, 2036z, 2037a, 2037b, 2037c, 2037d, 2037e, 2037f, 2037g, 2037h, 2037i, 2037j, 2037k, 2037l, 2037m, 2037n, 2037o, 2037p, 2037q, 2037r, 2037s, 2037t, 2037u, 2037v, 2037w, 2037x, 2037y, 2037z, 2038a, 2038b, 2038c, 2038d, 2038e, 2038f, 2038g, 2038h, 2038i, 2038j, 2038k, 2038l, 2038m, 2038n, 2038o, 2038p, 2038q, 2038r, 2038s, 2038t, 2038u, 2038v, 2038w, 2038x, 2038y, 2038z, 2039a, 2039b, 2039c, 2039d, 2039e, 2039f, 2039g, 2039h, 2039i, 2039j, 2039k, 2039l, 2039m, 2039n, 2039o, 2039p, 2039q, 2039r, 2039s, 2039t, 2039u, 2039v, 2039w, 2039x, 2039y, 2039z, 2040a, 2040b, 2040c, 2040d, 2040e, 2040f, 2040g, 2040h, 2040i, 2040j, 2040k, 2040l, 2040m,

- Geological Survey Water-Supply Paper 2175, 284 p.
- Rantz, S.E., and Others, 1982b, Measurement and computation of streamflow: Volume 2, Computation of discharge: Reston, VA, U.S. Geological Survey Water-Supply Paper 2175, 631 p.
- Redmond, K.T., Enzel, Yehouda, House, P.K., and Biondi, Franco, 2002, Climate variability and flood frequency at decadal to millennial time scales, *in* House, P.K., Webb, R.H., Baker, V.R., and Levish, D.R., *Water Science and Application Series*, v. 5, *Ancient Floods, Modern Hazards*, American Geophysical Union, 21–45 p., accessed August 30, 2017, at <https://doi.org/10.1029/WS005p0021>.
- Reed, Duncan, 1999, *Flood Estimation Handbook, Volume 1: Overview*, Institute of Hydrology, Wallingford, Oxfordshire, UK, 108 p.
- Reis, D.S., Stedinger, J.R., and Martins, E.S., 2005, Bayesian GLS regression with application to LP3 regional skew estimation: *Water Resources Research*, v. 41, p. W10,419.
- Ries, K.G., Guthrie, J.G., Rea, A.H., Steeves, P.A., and Stewart, D.W., 2008, *StreamStats: A Water Resources Web Application*, U.S. Geological Survey, 6 p.
- Rogger, M., Kohl, B., Pirkel, H., Viglione, A., Komma, J., Kirnbauer, R., Merz, R., and Blöschl, G., 2012, Runoff models and flood frequency statistics for design flood estimation in Austria – Do they tell a consistent story?: *Journal of Hydrology*, v. 456–457, p. 30–43, accessed August 30, 2017, at <https://doi.org/10.1016/j.jhydrol.2012.05.068>.
- Rosner, B., 1983, Percentage points for a generalized ESD many outlier procedure: *Technometrics*, v. 25, no. 2, p. 165–172.
- Russell, S.O., 1982, Flood probability estimation: *Journal of Hydraulic Engineering*, v. 108, no. HY1, p. 63–73.
- Salas, J., and Obeysekera, J., 2014, Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events: *Journal of Hydrologic Engineering*, v. 19, no. 3, p. 554–568, accessed August 30, 2017, at [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000820](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000820).
- Salas, J.D., 1993, Analysis and Modeling of Hydrologic Time Series, *in* Maidment, D.R., *Handbook of Hydrology*, McGraw Hill, Inc., 19.1–19.72 p.
- Salas, J.D., and Boes, D.C., 1980, Shifting level modelling of hydrologic series: *Advances in Water Resources*, v. 3, no. 2, p. 59 – 63, accessed August 30, 2017, at [https://doi.org/10.1016/0309-1708\(80\)90028-7](https://doi.org/10.1016/0309-1708(80)90028-7).
- Salas, J.D., Wohl, E.E., and Jarrett, R.D., 1994, Determination of flood characteristics using systematic, historical and paleoflood data, *in* *Coping With Floods*, edited by G. Rossi, N. Harmancioglu, and V. Yevjevich, Kluwer Academic Publishers, Netherlands, p. 111–134.
- Sanders, C.L., Kubik, H.E., Hoke, J.T., and Kirby, W.H., 1990, Flood frequency of the Savannah River at Augusta, Georgia, U.S. Geological Survey, Reston, VA, 87 p.
- Sando, S.K., Driscoll, D.G., and Parrett, Charles, 2008, Peak-flow frequency estimates based on data through water year 2001 for selected streamflow-gaging stations in South Dakota, U.S. Geological Survey, Reston, VA, 367 p.
- Sauer, V.B., and Meyer, R.W., 1992, Determination of error in individual discharge measurements, U.S. Geological Survey, 21 p.
- Sauer, V.B., and Turnipseed, D.P., 2010, Stage measurement at gaging stations, U.S. Geological Survey, 45 p.
- Sauer, V.B., Thomas, W.O., Stricker, V.A., and Wilson, K.V., 1983, Flood characteristics of urban watersheds in the United States, U.S. Geological Survey, Reston, VA, 63 p.
- Sigafoos, R.S., 1964, Botanical evidence of floods and flood-plain deposition, U.S. Geological Survey, 35 p.
- Singh, V.P., 1995, *Computer Models of Watershed Hydrology: Highlands Ranch, CO*, Water Resources Publications, 1130 p.

DRAFT: August 31, 2017

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

- Smith, J.A., Villarini, Gabriele, and Baeck, M.L., 2011, Mixture distributions and the hydroclimatology of extreme rainfall and flooding in the Eastern United States: *Journal of Hydrometeorology*, v. 12, p. 294–309, accessed August 30, 2017, at <https://doi.org/10.1175/2010JHM1242.1>.
- Southard, R.E., and Veilleux, A.G., 2014, Methods for estimating annual exceedance-probability discharges and largest recorded floods for unregulated streams in rural Missouri, U.S. Geological Survey, Reston, VA, 39 p.
- Stedinger, J.R., and Cohn, T.A., 1986, Flood frequency analysis with historical and paleoflood information: *Water Resources Research*, v. 22, no. 5, p. 785–793.
- Stedinger, J.R., and Cohn, T.A., 1987, Historical flood-frequency data: Its value and use, *in* Regional Flood Frequency Analysis, edited by V. P. Singh, D. Reidel, p. 273–286.
- Stedinger, J.R., and Griffis, V.W., 2008, Flood frequency analysis in the United States: Time to update: *Journal of Hydrologic Engineering*, v. 13, no. 4, p. 199–204.
- Stedinger, J.R., and Griffis, V.W., 2011, Getting from here to where? Flood frequency analysis and climate: *JAWRA Journal of the American Water Resources Association*, v. 47, no. 3, p. 506–513, accessed August 30, 2017, at <https://doi.org/10.1111/j.1752-1688.2011.00545.x>.
- Stedinger, J.R., Surani, R., and Therivel, R., 1988, Max user's guide: A program for flood frequency analysis using systematic-record, historical, botanical, physical paleohydrologic and regional hydrologic information using maximum likelihood techniques, Cornell University, 111 p.
- Stedinger, J.R., Vogel, R.M., and Foufoula-Georgiou, Efi, 1993, Frequency analysis of extreme events, *in* Maidment, D.R., *Handbook of Hydrology*, McGraw Hill, Inc., 18.1-18.66 p.
- Stewart, J.E., and Bodhaine, G.L., 1961, Floods in the Skagit River basin, Washington, U.S. Geological Survey, Reston, VA, 66 p.
- Sutley, D.E., Klinger, R.E., Bauer, T.R., and Godaire, J.E., 2008, Trinity Dam detailed hydrologic hazard analysis using the Stochastic Event Flood Model, U.S. Department of Interior, Bureau of Reclamation, Denver, CO, 84 p.
- Suttie, R.H., 1928, Report on the Water Resources of Connecticut, Connecticut Geol. and Nat. Hist. Survey, Hartford, CT, 168 p.
- Sveinsson, O. G.B., Salas, J.D., Boes, D.C., and Sr, R. A.P., 2003, Modeling the dynamics of long-term variability of hydroclimatic processes: *Journal of Hydrometeorology*, v. 4, p. 489–505, accessed August 30, 2017, at [https://doi.org/10.1175/1525-7541\(2003\)004\(0489:MTDOLV\)2.0.CO;2](https://doi.org/10.1175/1525-7541(2003)004(0489:MTDOLV)2.0.CO;2).
- Swain, R.E., England Jr, J.F., Bullard, K.L., and Raff, D.A., 2006, Guidelines for Evaluating Hydrologic Hazards, U.S. Bureau of Reclamation, Denver, CO, 83 p.
- Tasker, G.D., 1978, Flood frequency analysis with a generalized skew coefficient: *Water Resources Research*, v. 14, no. 2, p. 373–376.
- Tasker, G.D., 1983, Effective record length for the T-year event: *Journal of Hydrology*, v. 64, p. 39–47.
- Tasker, G.D., and Stedinger, J.R., 1986, Regional skew with weighted LS regression: *Journal of Water Resources Planning and Management*, v. 112, no. 2, p. 225–237.
- Tasker, G.D., and Stedinger, J.R., 1989, An operational GLS model for hydrologic regression: *Journal of Hydrology*, v. 111, no. 4, p. 361–375.
- Thomas, W.O., 1982, An evaluation of flood frequency estimates based on rainfall/runoff modeling: *JAWRA Journal of the American Water Resources Association*, v. 18, no. 2, p. 221–229, accessed August 30, 2017, at <https://doi.org/10.1111/j.1752-1688.1982.tb03964.x>.
- Thomas, W.O., 1985, A uniform technique for flood frequency analysis: *Journal of Water Resources Planning and Management*, v. 111, no. 3, p. 321–337.
- Thomas, W.O., 1987, Comparison of flood-frequency estimates based on observed and model-generated peak flows, *in* *Hydrologic Frequency Modeling*,

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

- 1 edited by V. P. Singh, Louisiana State University, 43  
2 Baton Rouge, D. Reidel Publishing Company, p. 44  
3 149–161. 45
- 4 Thomas, W.O., Grimm, M.M., and McCuen, R.H., 46  
5 2001, Evaluation of Flood Frequency Estimates 47  
6 for Ungaged Watersheds, Hydrologic Frequency 48  
7 Work Group, Subcommittee on Hydrology. 49
- 8 Thomson, M.T., Gannon, W.B., Thomas, M.P., and 50  
9 Hayes, G.S., 1964, Historical floods in New Eng- 51  
10 land, U.S. Geological Survey, Reston, VA, 105 p. 52
- 11 Turnipseed, D.P., and Sauer, V.B., 2010, Discharge 53  
12 measurement at gaging stations, U.S. Geological 54  
13 Survey, 87 p. 55
- 14 U.S. Army Corps of Engineers, 1973, Storm Rain- 56  
15 fall in the United States, 1945 - 1973, U.S. Army 57  
16 Corps of Engineers, Washington, DC. 58
- 17 U.S. Army Corps of Engineers, 1982, Mixed- 59  
18 Population Frequency Analysis, U.S. Army Corps 60  
19 of Engineers, Hydrologic Engineering Center, 61  
20 43 p. 62
- 21 U.S. Army Corps of Engineers, 1993, Hydrologic 63  
22 Frequency Analysis, U.S. Army Corps of Engi- 64  
23 neers. 65
- 24 U.S. Army Corps of Engineers, 2008, Orestimba 66  
25 Creek Hydrology: A Reevaluation Based on 67  
26 Updated Peak and Volume Frequency Curves, U.S. 68  
27 Army Corps of Engineers, Water Management 69  
28 Section, Sacramento District. 70
- 29 U.S. Bureau of Reclamation and Utah State Uni- 71  
30 versity, 1999, A Framework for Characterizing 72  
31 Extreme Floods for Dam Safety Risk Assessment, 73  
32 U.S. Bureau of Reclamation, 67 p. 74
- 33 USWRC, 1967, A uniform technique for deter- 75  
34 mining flood flow frequencies, Bulletin No. 15, 76  
35 U.S. Water Resources Council, Subcommittee on 77  
36 Hydrology, Washington, D.C. 78
- 37 USWRC, 1976, Guidelines for Determining Flood 79  
38 Flow Frequency, Bulletin No. 17, U.S. Water 80  
39 Resources Council, Subcommittee on Hydrology, 81  
40 Washington, D.C. 82
- 41 Veilleux, Andrea, 2011, Bayesian GLS Regression 83  
42 For Regionalization Of Hydrologic Statistics, 84  
43 Floods And Bulletin 17 Skew, Ph.D. dissertation, 85  
44 Cornell University.
- 45 Veilleux, A.G., Stedinger, J.R., and Lamontagne, 86  
46 J.R., 2011, Bayesian WLS/GLS Regression for 87  
47 Regional Skewness Analysis for Regions with 88  
48 Large Cross-Correlations among Flood Flows, 89  
49 *in* World Environmental and Water Resources 90  
50 Congress 2011 Bearing Knowledge for Sustain- 91  
51 ability, ASCE.
- 52 Viglione, A., Merz, R., Salinas, J.L., and Blöschl, 92  
53 G., 2013, Flood frequency hydrology: 3. A 93  
54 Bayesian analysis: Water Resources Research, 94  
55 v. 49, no. 2, p. 675–692, accessed August 30, 2017, 95  
56 at <https://doi.org/10.1029/2011WR010782>.
- 57 Villarini, Gabriele, Serinaldi, Francesco, Smith, J.A., 96  
58 and Krajewski, W.F., 2009a, On the stationarity 97  
59 of annual flood peaks in the continental United 98  
60 States during the 20th century: Water Resources 99  
61 Research, v. 45, no. 8, accessed August 30, 2017, 100  
62 at <https://doi.org/10.1029/2008WR007645>.
- 63 Villarini, Gabriele, Smith, J.A., Serinaldi, Francesco, 101  
64 Bales, Jerad, Bates, P.D., and Krajewski, W.F., 102  
65 2009b, Flood frequency analysis for nonstation- 103  
66 ary annual peak records in an urban drainage 104  
67 basin: Advances in Water Resources, v. 32, no. 8, 105  
68 p. 1255 – 1266, accessed August 30, 2017, at 106  
69 <https://doi.org/10.1016/j.advwatres.2009.05.003>.
- 70 Vogel, R.M., and Kroll, C.N., 1991, The value of 107  
71 streamflow record augmentation procedures in 108  
72 low-flow and flood-flow frequency analysis: Jour- 109  
73 nal of Hydrology, v. 125, no. 34, p. 259 – 276, 110  
74 accessed August 30, 2017, at [https://doi.org/10.1016/0022-1694\(91\)90032-D](https://doi.org/10.1016/0022-1694(91)90032-D). 111
- 75 Vogel, R.M., and Stedinger, J.R., 1984, Flood-Plain 112  
76 Delineation in Ice Jam Prone Regions: Journal of 113  
77 Water Resources Planning and Management, v. 114  
78 110, no. 2, p. 206–219. 115
- 79 Vogel, R.M., and Stedinger, J.R., 1985, Minimum 116  
80 variance streamflow record augmentation proce- 117  
81 dures: Water Resources Research, v. 21, no. 5, 118  
82 p. 715–723. 119
- 83 Vogel, R.M., Matalas, N.C., England, J.F., and 120  
84 Castellarin, Attilio, 2007, An assessment of 121  
85

DRAFT: August 31, 2017

PROVISIONAL  
THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.



- exceedance probabilities of envelope curves:  
Water Resources Research, v. 43, no. 7, accessed  
August 30, 2017, at <https://doi.org/10.1029/2006WR005586>, w07403.
- Wahl, K.L., Thomas, W.O., and Hirsch, R.M., 1995,  
The stream-gaging program of the U.S. Geological  
Survey: Reston, VA, U.S. Geological Survey  
Circular 1123, 22 p.
- Weaver, J.C., Feaster, T.D., and Gotvald, A.J., 2009,  
Magnitude and Frequency of Rural Floods in  
the Southeastern United States, 2006: Volume  
2, North Carolina, U.S. Geological Survey, Reston,  
VA, 111 p.
- Webb, R.H., and Betancourt, J.L., 1992, Climatic  
variability and flood frequency of the Santa Cruz  
River, Pima County, Arizona: Reston, VA, U.S.  
Geological Survey Water-Supply Paper 2379,  
40 p.
- Webb, R.H., and Jarrett, R.D., 2002, One-  
dimensional estimation techniques for discharges  
of paleofloods and historical floods, *in* House,  
P.K., Webb, R.H., Baker, V.R., and Levish, D.R.,  
Water Science and Application Series, v. 5,  
Ancient Floods, Modern Hazards, American Geo-  
physical Union, 111–125 p., accessed August 29,  
2017, at <https://doi.org/10.1029/WS005p0111>.  
Wiley, J.B., and Atkins, J. T.J., 2010, Estimation of  
Flood-Frequency Discharges for Rural, Unregu-  
lated Streams in West Virginia, U.S. Geological  
Survey, Reston, VA, 78 p.  
Wohl, E.E., and Enzel, Y., 1995, Data for palaeohy-  
drology, *in* Gregory, K.J., Starkel, L., and Baker,  
V.R., Global Continental Palaeohydrology, John  
Wiley & Sons, Ltd., 24-59 p.  
Yanosky, T.M., 1983, Evidence of floods on the  
potomac river from anatomical abnormalities in  
the wood of flood-plain trees, U.S. Geological  
Survey, 42 p.  
Yanosky, T.M., and Jarrett, R.D., 2002, Den-  
drochronologic evidence for the frequency and  
magnitude of paleofloods, *in* House, P.K., Webb,  
R.H., Baker, V.R., and Levish, D.R., Water  
Science and Application Series, v. 5, Ancient  
Floods, Modern Hazards, American Geophys-  
ical Union, 77–89 p., accessed August 29, 2017, at  
<https://doi.org/10.1029/WS005p0077>.  
Yen, B.C., 1970, Risks in hydrologic design of engi-  
neering projects: Journal of the Hydraulics Divi-  
sion, ASCE, v. 96, no. HY4, p. 959–966.

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017



DRAFT: August 31, 2017

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

## Glossary

- annual exceedance probability (AEP)** — The probability that flooding will occur in any given year considering the full range of possible annual floods.
- annual flood** — The highest instantaneous peak discharge in each year of record. Practically, this is the highest value observed in the record of 15 minute or 60 minute values, depending on the recording interval of the device. Sometimes the maximum mean daily discharge is used on larger rivers.
- annual flood series** — A list of annual maximum floods.
- annual series** — A general term for a set of any kind of data in which each item is the maximum or minimum in a year.
- autocorrelation** — The presence of autocorrelation indicates that the data in the time series are not random. Rather, future values are correlated with past values. Autocorrelation is calculated as the correlation between the values in a time series and the values in that same time series lagged by one or more time steps (that is, the correlation between  $X_i$  and  $X_{i+k}$  where  $i$  is the time step and  $k$  is the lag). Also known as serial correlation.
- base discharge (for peak discharge)** — A discharge value, determined for selected stations, above which peak discharge data are published. The base discharge at each station is selected so that an average of about three peak flows per year will be published (Langbein and Iseri, 1960).
- binomial censored data** — Floods that exceeded a threshold, where one knows only that a flood was larger than some level, but does not know the magnitude of the flood (Russell, 1982; Stedinger and Cohn, 1986).
- broken record** — A systematic record that is divided into separate continuous segments because of deliberate discontinuation of recording for significant periods of time. This typically occurs when a gage is shut off due to funding, prioritization, or other hydrological or management reasons, then reestablished at a later time (several years, rather than weeks or months) at the same location. *See also discontinued record.*
- censored data** — In a sample size of  $n$ , a known number of observations is missing at either end or at both ends (David, 1981; Cohen, 1991).
- coefficient of skewness** — A numerical measure or index of the lack of symmetry in a frequency distribution. Function of the third moment of magnitudes about their mean, a measure of asymmetry. Also known as coefficient of skew or **skew coefficient**.
- confidence limits** — Computed values on both sides of an estimate of a parameter or quantile that show for a specified probability the range in which the true value of the parameter or quantile lies.
- crest-stage gage (CSG)** — A simple, economical, reliable, and easily installed device for obtaining the elevation of the flood crest of streams (Sauer and Turnipseed, 2010). These gages are nonrecording and consist of a partial streamflow record. Flow intervals and perception thresholds are needed to describe each year of the flood record.

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

## 54 Guidelines for Determining Flood Flow Frequency – Bulletin 17C

**cross correlation** — A measure of similarity, interdependence, or relationship between two time series of observations in space at the same point or lagged points in time.

**discontinued record** — A systematic record where streamflow data collection and measurements have ceased. This typically occurs when a gage is shut off due to funding, prioritization, other hydrological or management reasons, and not reestablished. *See also* **broken record**.

**exceedance** — Knowledge that the magnitude (discharge or stage) of a flood was larger than some level or threshold. For example, the flood exceeded the bridge deck.

**exceedance frequency** — The percentage of values that exceed a specified magnitude, 100 times exceedance probability.

**exceedance probability** — Probability that a random event will exceed a specified magnitude in a given time period, usually one year unless otherwise indicated.

**Expected Moments Algorithm (EMA)** — A generalized method of moments procedure to estimate the P-III distribution parameters using the entire dataset, simultaneously employing regional skew information and a wide range of historical flood and threshold-exceedance information, while adjusting for any potentially influential low floods, missing values from an incomplete record, or zero flood years.

**extraordinary flood** — Those floods that are the largest magnitude at a gaging station or miscellaneous site and that substantially exceed the other flood observations (Costa and Jarrett, 2008).

**gage base** — The minimum stage or discharge level at a gaging station, below which observations are not recorded or published. Also called **base discharge**.

**gaging record** — Streamflow data collected at streamflow-gaging stations. A gaging record can consist of systematic data and historical flood data.

**gaging station** — A selected site on a stream equipped and operated to furnish basic data from which continuous, systematic records of stage and discharge may be obtained (Grover and Harrington, 1943; Rantz and Others, 1982).

**generalized skew coefficient** — *See* **regional skew coefficient**.

**high-water mark (HWM)** — Typically recent (hours to weeks) physical evidence of the (approximate) maximum flood stage (Jarrett and England, 2002). The physical evidence generally is of three types: (1) deposits along channel margins and in vegetation that consist of very light, floatable material such as pine needles, seeds, small twigs, grasses, and very fine sediments; (2) damage to vegetation such as bent or matted grasses, twigs, and branches, stripped leaves or bark; and (3) small erosional features such as scour lines. Benson and Dalrymple (1967) and Koenig and others (2016) discuss identification, preservation, and rating of high-water marks. The HWM evidence is typically short-lived (weeks), but woody debris may last from several years to several decades in arid and semi-arid climates (Baker, 1987), and geomorphic evidence can be preserved for millennia. Physical evidence, such as marks on buildings and other structures, is also long lived. This stage may represent a maximum discharge when a single-valued relationship exists between stage and discharge; Costa and Jarrett (2008) describe other hydraulic situations. *See also* **paleostage indicator**.

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

- historical data** — Broad category of data collected by humans prior to establishing systematic protocols; it generally consists of diaries, written accounts of settlements, folklore, and descriptions that may document periods where extreme weather and (or) floods have occurred. It may also be used to infer times when there have been no large floods. These accounts were recorded in a manner that was preserved well enough that we know about it today.
- historical floods** — Flood events which were directly observed by humans, generally in a nonsystematic manner by nonhydrologists (Baker, 1987). These events usually occurred and were described in some qualitative and (or) quantitative fashion prior to the systematic record. Information about the floods was recorded and preserved well enough so that we know about it today.
- homogeneity** — Records from the same populations. Floods may be from different populations because they occurred before the building of a dam and after the building of a dam, or before the watershed was urbanized and after it became urbanized, or because some are generated by summer storms and others by snowmelt, or because some were generated in El Niño years and some were in other years. It may be difficult in some cases to definitively say if the flood record is homogeneous.
- incomplete record** — A streamflow record in which some peak flows are missing because they were too low or high to record or the gage was out of operation for a short period because of flooding.
- interval data** — Floods whose magnitude are not known exactly, but are known to fall within a range or interval (Stedinger and others, 1988; Cohn and others, 1997).
- level of significance** — The probability of rejecting a hypothesis when it is in fact true. At a “10-percent” level of significance, the probability is 1/10.
- low outlier** — *See outlier.*
- mean-square error** — Sum of the squared differences between the true and estimated values of a quantity divided by the number of observations. It can also be defined as the bias squared plus the variance of the quantity (Stedinger and others, 1993).
- method of moments** — A standard statistical computation for estimating the parameters of a distribution from the moments of the sample data.
- Multiple Grubbs-Beck Test (MGBT)** — A statistical test used to identify multiple potentially influential low flood observations in an annual maximum time series.
- nonexceedance** — Knowledge that the magnitude (discharge or stage) of a flood was less than some level or threshold.
- outlier** — Outliers (extreme events) are observations that are exceedingly low or high compared to the distributional properties of the vast majority of the data. When plotted, along with a reasonably fitted CDF to the data, the outlier values plot far from the fitted line at the low or high ends of the distribution. A CDF, such as the LP-III, may not fit datasets with outliers, and the fitted curve usually fails to fit the bulk of the data as well as the outliers. Low outliers are outliers at the low end of the dataset, near zero, at least in comparison with the rest of the data. On a log-probability plot, the low outliers impart a strong downward curvature and a downward-drooping lower tail to the frequency curve. In comparison with the lower tail, the upper tail of the low-outlier-affected curve may appear relatively flat.

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

**paleoflood data** — Physical evidence of past floods and their ages as observed from the geologic record or from botanical evidence. Paleoflood data typically consist of observations on individual past floods such as those derived from slackwater deposits, boulder bars, silt lines, or botanical information, that are collected as part of a paleoflood hydrology study (Benito and O'Connor, 2013). It can also consist of periods of landscape stability that can be used to place limits on flood magnitude over time, such as paleohydrologic bounds (Levish, 2002). Paleoflood data are distinguished from historical flood data, as a separate line of evidence, by the use of applied field geology techniques to examine and describe the geomorphic and stratigraphic context of extreme floods. In some cases, there is overlap between historical and paleoflood data, as historical and cultural artifacts such as barbed wire, beer cans (House and Baker, 2001), or pottery may be observed and used in dating and estimation of floods.

**paleoflood hydrology** — The study of past or ancient floods that occurred prior to the time of human observation or direct measurement by modern hydrologic procedures (Baker, 1987). Paleoflood hydrology has also been defined as “the study of the movements of water and sediment in channels before the time of continuous hydrologic records or direct measurements” (Costa, 1986).

**paleohydrologic bound** — A time interval during which a given discharge has not been exceeded (Levish, 2002). The term is sometimes shortened to “bound”. The paleohydrologic bound represents stages and discharges that have not been exceeded since the geomorphic surface stabilized. Bounds are appropriate for paleohydrologic information and are not dependent on human observation of a particular event, but on the physical setting (hydraulic and geomorphic). Also known as “paleoflood bound”.

**paleostage indicator (PSI)** — An erosional or depositional feature that recorded the near peak stage of an individual flood (Jarrett and England, 2002) prior to human observation. Indirect evidence of the stage of past floods includes botanical evidence and sedimentological deposits (Jarrett, 1991). Large floods, especially in high-gradient channels, can transport and deposit coarse material (gravel, boulders, and woody debris, and so forth) that may be interpreted as HWMs. The primary differences between PSIs and HWMs are: (1) HWMs represent events that occurred “more recent” in time due to their relatively short preservation length as compared to PSIs; and (2) some PSIs may not represent the exact peak stage.

**Partial-Duration Series (PDS)** — A list of all flows (such as flood peaks) that exceed a chosen base stage or discharge, regardless of the number of peaks occurring in a year. Also called basic-stage flood series, or floods above a base (Langbein and Iseri, 1960).

**percent chance** — A probability multiplied by 100.

**perception threshold** — The stage or flow above which it is estimated a source would provide information on the flood peak in any given year. Perception thresholds ( $T_{y,lower}$ ;  $T_{y,upper}$ ) reflect the range of flows that would have been measured or recorded had they occurred. If an event magnitude had occurred in a specific year, there is information that would indicate it would have been “recorded” in a manner that we could perceive it today. Perception thresholds describe the range of measurable potential discharges and are independent of the actual peak discharges that have occurred. They are used to provide a rank and record length for each reported flood peak (Gerard and Karpuk, 1979). Perception thresholds are used for historical data, when the information provided is based on human observation. They are also used to describe a paleoflood period and paleoflood data. In addition, perception thresholds are used to properly accommodate unrecorded floods below a “gage base”. A perception threshold is allocated to each information source for each year  $Y$  of the flood record. Perception thresholds may involve a significant amount of judgment on the part of the scientist and (or) historian regarding, for any given year, what is

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017



- the smallest event that would have been recorded (in a physical or textural manner) such that we would actually know about it today Also known as **threshold**.
- population** — The entire (usually infinite) number of data from which a sample is taken or collected. The total number of past, present, and future floods at a location on a river is the population of floods for that location, even if the floods are not measured or recorded.
- potentially influential low flood (PILF)** — In an annual maximum flood series, small-magnitude flows (including zeros) that do not represent the physical processes that cause the largest flood observations. These PILFs can exert high leverage and influence on the flood frequency distribution.
- quantile** — Estimate of the flood magnitude  $Q$  for exceedance probability  $p$  from a fitted distribution.
- record augmentation** — A procedure to improve the accuracy of the moments (mean and variance) of a short-record flood series by using information from longer records at nearby locations with high cross correlation (Matalas and Jacobs, 1964; Stedinger and others, 1993).
- record extension** — The creation of a longer flood-flow record (individual floods) at a site with a short record, by using flood observations at a long-record site with high cross-correlation. The technique can also be used to fill in missing observations (Hirsch and others, 1993; Stedinger and others, 1993).
- regional skew coefficient** — A skew coefficient derived by a procedure that integrates values obtained at many locations.
- robustness** — In flood frequency, a procedure that is reasonably efficient when the assumed characteristics of the flood distribution are true, while not doing poorly when those assumptions are violated (Kuczera, 1982; Coia and others, 1993).
- sample** — An element, part, or fragment of a “population.” Every hydrologic record is a sample of a much longer record.
- serial correlation** — See **autocorrelation**.
- skew coefficient** — See **coefficient of skewness**.
- standard deviation** — A measure of the dispersion or precision, of a series of statistical values such as precipitation or streamflow. It is the square root of the sum of squares of the deviations from the arithmetic mean divided by the number of values or events in the series. It is standard practice to divide by the number of values minus one in order to get an unbiased estimate of the variance from the sample data.
- standard error** — An estimate of the standard deviation of a statistic. Often calculated from a single set of observations. Calculated like the standard deviation but differing from it in meaning.
- systematic data** — Data that are collected at regular, prescribed intervals under a defined protocol. In the context of streamflow, systematic data consist of discharge and stage data collected at regular, prescribed intervals, typically at gaging stations. Synonymous with systematic record.
- threshold** — See **perception threshold**.
- variance** — A measure of the amount of spread or dispersion of a set of values around their mean, obtained by calculating the mean value of the squares of the deviations from the mean, and, hence, equal to the square of the standard deviation.

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

## 58 Guidelines for Determining Flood Flow Frequency – Bulletin 17C

1 **weighted means** — A value obtained by multiplying each of a series of values by its assigned weight and  
2 dividing the sum of those products by the sum of the weights.

DRAFT: August 31, 2017

---PROVISIONAL---  
THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.  
IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).  
IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

## APPENDIXES

1

DRAFT: August 31, 2017

---PROVISIONAL---  
THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.  
IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).  
IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

## Appendix 1. List of Symbols

---

1

2

DRAFT: August 31, 2017

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.



DRAFT: August 31, 2017

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

## List of Symbols

1

The following is a list of symbols that are used in this report.

2

DRAFT: August 31, 2017

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

Table 1.1. List of symbols used in this report.

Symbol	Description
$a$	Plotting position parameter that is dependent on an assumed distribution ( $0 < a < 0.5$ ); $a = 0 \equiv$ Weibull; $a = 0.5 \equiv$ Hazen (eq. 2)
$e_s$	Number of floods/records that exceed a censoring level, such as a historical threshold $T_h$ or low-outlier threshold $T_{PILF}$ , during the systematic record $n_s$ ( $e_s \leq k; e_s < n_s$ )
$e_h$	Number of floods/records that exceed the historical threshold $T_h$ during the historical/paleoflood period $n_h$ ( $e_h \leq k; e_h < n_h$ )
$g$	Total number of known flood (observations) during the entire period of observation record $n$ ( $g = n_s + k - e_s = n_s + e_h$ )
$\hat{\gamma}$	At-site (station) sample skew coefficient (in log space)
$G$	Regional sample skew coefficient (in log space)
$\tilde{G}$	Weighted skew coefficient
$k$	Total number of floods/records that exceed a censoring level, such as a historical threshold $T_h$ or low-outlier threshold $T_{PILF}$ , during the entire period of observation record $n$ ( $k = e_s + e_h$ )
$\hat{\mu}$	At-site (station) sample mean (in log space)
$\hat{\sigma}$	At-site (station) sample standard deviation (in log space)
$n$	Total peak-flow period of record (years), including systematic $n_s$ and historical $n_h$ periods, as available, where $n = n_s + n_h$
$n_h$	Length of the historical period (years); possibly includes a paleoflood period ( $n_h < n$ )
$n_s$	Length of the peak-flow systematic (gaging) record (years) ( $n_s \leq n$ )
$p$	Annual exceedance probability (AEP), $p = 1 - q$
$q$	Cumulative probability, $q = 1 - p$
$Q$	Flood discharge
$Q_b$	Base discharge. Can be a constant, or vary with each year at a gaging station or CSG
$Q_p$	Discharge quantile for annual exceedance probability $p$
$Q_q$	Discharge quantile for cumulative probability $q$ , equivalent to $Q_p$
$Q_Y$	Flood discharge estimate in year $Y$
$Q_{Y,lower}$	Discharge lower bound for year $Y$ in EMA

DRAFT August 31, 2017

---PROVISIONAL---  
 THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
 PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
 PEER REVIEW PLAN.  
 IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
 GEOLOGICAL SURVEY (USGS).  
 IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
 TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

## Appendix 2. Subcommittee and Work Group Members

---

1

2

DRAFT: August 31, 2017

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.



## Subcommittee and Work Group Members

The Subcommittee on Hydrology (SOH) is a ~~sub-group~~ subgroup under the Advisory Committee on Water Information (ACWI). The purpose of the SOH is to improve the availability and reliability of surface-water quantity information needed for hazard mitigation, water supply and demand management, and environmental protection. The SOH coordinates and oversees technical working groups, including the Hydrologic Frequency Analysis Work Group (HFAWG). The SOH sponsored this HFAWG work effort to prepare the update to these Guidelines. Current SOH membership is listed in ~~Table~~ table 2.1. Further details about SOH and its activities are available at <https://acwi.gov/hydrology/index.html>.

The overall goal of the Hydrologic Frequency Analysis Work Group is to recommend procedures to increase the usefulness of the current guidelines for Hydrologic Frequency Analysis computations (~~e.g.~~ for example, Bulletin 17B) and to evaluate other procedures for frequency analysis of hydrologic phenomena. The work group forwards draft papers and recommendations to the Subcommittee on Hydrology of ACWI for appropriate action. As part of these activities, the HFAWG oversaw the revision to these Guidelines. Current HFAWG membership is listed in ~~Table~~ table 2.2. Further details about HFAWG and its activities are available at <https://acwi.gov/hydrology/Frequency/index.html>.

**Table 2.1.** Subcommittee on Hydrology ~~Members~~ members.

Member Organization	Representative
Association of State Floodplain Managers	Wilbert O. Thomas, Jr.
BECKER	Martin Becker
DOI/Bureau of Land Management	Robert Boyd
DOI/Bureau of Reclamation	Dr. Ian Ferguson
DOI/Office of Surface Mining	TBD
DOI/US Geological Survey	Robert Mason ( <del>Chair</del> ) <u>(Chair)</u>
Federal Energy Regulatory Commission	Dr. S. Samuel Lin
Federal Highway Administration	Brian Beucler
Global Ecosystems Center	Don Woodward
NASA/Goddard Space Flight Center	Sujay Kumar
National Hydrologic Warning Council	Ben Pratt
National Science Foundation	Dr. Thomas Torgersen
NOAA/National Weather Service	Victor Hom
US Army Corps of Engineers	Dr. Chandra Pathak
USDA/Agricultural Research Service (ARS)	Dr. David C. Goodrich
USDA/Natural Resources Conservation Service (NRCS)	Claudia Hoeft
USDA/U.S. Forest Service	Michael Eberle
U.S. Environmental Protection Agency	Pravin Rana
USDHS/Federal Emergency Management Agency (FEMA)	Dr. Siamak Esfandiary ( <del>Vice Chair</del> ) <u>(Vice Chair)</u>
U.S. Nuclear Regulatory Commission (NRC)	Thomas J. Nicholson

~~PROVISIONAL~~

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

**Table 2.2.** Hydrologic Frequency Analysis Work Group ~~Members~~members.

Member Name	Organization	Location
Wilbert O. Thomas <del>Jr. (Chair)</del> , Jr. (Chair)	Michael Baker International	Manassas, VA
Dr. Siamak Esfandiary	Federal Emergency Management Agency	Crystal City, VA
Don Woodward	Global Ecosystems Center	Derwood, MD
Martin Becker	BECKER	Atlanta, GA
Dr. Timothy Cohn	U.S. Geological Survey	Reston, VA
Dr. Beth Faber	U.S. Army Corps of Engineers	Davis, CA
Dr. John England	U.S. Army Corps of Engineers	Lakewood, CO
Prof. Jerry Stedinger	Cornell University	Ithaca, NY
Dr. Zhida Song-James	Consulting Hydrologist	Fairfax, VA
Dr. Jerry Coffey	Mathematical Statistician	Middletown, VA
Joe Krolak	Federal Highway Administration	Washington, DC
William Merkel	Natural Resources Conservation Service	Beltsville, MD
Dr. Sanja Perica	National Weather Service	Silver Spring, MD
Thomas Nicholson	Nuclear Regulatory Commission	Rockville, MD
Dr. S. Samuel Lin	Federal Energy Regulatory Commission	Washington, DC
Mike Eiffe (through Sept. 2014)	Tennessee Valley Authority	Knoxville, TN
Curt Jawdy	Tennessee Valley Authority	Knoxville, TN

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

## Appendix 3. Data Sources and Representation

---

1

2

DRAFT: August 31, 2017

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

## Data Sources

This appendix provides some representative data sources for flood frequency. Systematic records, historical data, paleoflood data and botanical information, regional information, and precipitation and climate information are briefly described. These sources are intended to be used as references and starting points for data collection, and are not all ~~-inclusive~~inclusive. Available sources and some websites for these data can be found at <https://acwi.gov/hydrology/Frequency/b17c/>.

## Systematic Records

Systematic records that may be useful for estimating flood frequency include the following: peak flows, daily flows, reservoir inflows and elevations, hydrograph data, and streamflow measurements. Annual maximum instantaneous peak streamflow and gage height data can be obtained from the USGS National Water Information System (NWIS) at <https://nwis.waterdata.usgs.gov/usa/nwis/peak>. Users need to be familiar with the NWIS qualification codes for peak streamflow and peak gage height at <https://nwis.waterdata.usgs.gov/nwis/peak?help>. Daily streamflow data can be obtained from various sources. The main data source is the USGS NWIS at [https://nwis.waterdata.usgs.gov/nwis/dv/?referred\\_module=sw](https://nwis.waterdata.usgs.gov/nwis/dv/?referred_module=sw). These data can be easily retrieved with software packages such as Hirsch and DeCicco (2015).

Other Federal agencies provide daily streamflow and extensive reservoir-related data, including elevations, inflows, and outflows. These data can be of direct use for extending discontinued streamflow gages and estimating unregulated flows.

The Bureau of Reclamation at <http://www.usbr.gov/> provides data through its five regions in the 17 ~~western~~ states Western States for numerous river locations and over 350 reservoirs. The Reclamation Hydromet databases provide data for the Great Plains Region and Pacific Northwest Region. Data within the Upper Colorado Region is obtained through reservoir operations at <http://www.usbr.gov/uc/>. Data within the Lower Colorado Region is obtained through river operations at <http://www.usbr.gov/lc/>. Reclamation's Mid-Pacific region provides data for many locations, including the Central Valley, through the California Data Exchange Center.

The U.S. Army Corps of Engineers provides streamflow and reservoir information, within the conterminous United States, through seven divisions. A map with links to each division is at <http://www.usace.army.mil>. Streamflow and reservoir data can be provided for specific projects or river basins, within each division. For example, the Northwestern Division provides data for the Missouri River basin through their reservoir control center.

Individual ~~state~~ State agencies provide streamflow information, typically through their Division of Water Resources or Division of Natural Resources. Some examples of streamflow databases by ~~states are~~ States are the following: California, Colorado, Oregon, and Minnesota. Local flood-control districts and organizations may also have relevant streamflow data.

Instantaneous data (15-minute data, unit values, and complete hydrographs), from 2007 to present for active streamgages, can be obtained from the USGS NWIS at [https://nwis.waterdata.usgs.gov/nwis/uv/?referred\\_module=sw](https://nwis.waterdata.usgs.gov/nwis/uv/?referred_module=sw). Hydrograph data from about the mid-1980s to 2007 can be obtained from the instantaneous data archive at <https://ida.water.usgs.gov/ida/>.

Data on manual measurements of streamflow and gage height, including indirect measurement, can be obtained from the USGS at <https://waterdata.usgs.gov/nwis/measurements>. These measurements are used to supplement and (or) verify the accuracy of the automatically recorded observations, as well as to compute streamflow based on gage height. They are valuable for flood frequency studies to aid hydrologists in understanding how the largest flood estimates are made (such as ~~an indirect~~ a slope-area indirect measurement), and in estimating uncertainty.

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017



## Historical Data

Historical flood data sources can be obtained from a variety of locations. This section describes some of those data sources useful for flood frequency, and is an excerpt from England (1998, Chapter 4) England (1998, chapter 4), updated with additional recent studies. A literature search is performed, followed by field studies and historical data collection efforts in the watershed and community of interest (Thomson and others, 1964; Aldridge and Hales, 1984).

One typically obtains ~~U.S. Geological Survey~~ USGS records as a first step in the search for historical information. Information on observed floods, occurring after about 1900, and that typically cause flooding of populated areas, damage, and sometimes deaths, are described in various USGS publications, such as Water-Supply Papers, Professional Papers, and Scientific Investigations Reports. The information generally consists of basin rainfall estimates, types of discharge or indirect measurements made, damage estimates, pictures of damaged structures, and erosion and deposition in channels and floodplains flood plains. In some cases, past historical flood dates, stages, and peak discharge estimates in the region are described in each report. For example, the report on the Arkansas River flood of June 3-5-5, 1921 (Follansbee and Jones, 1922) lists previous floods back to about 1844, based primarily on Denver and Rio Grande railroad records. Stewart and Bodhaine (1961) describe recent floods and present a historical flood chronology back to 1815 for the Skagit River basin Basin in Washington. In some cases, historic flood estimates are revised, such as the Skagit River near Concrete (Mastin, 2007). Many other Water-Supply Papers present historical flood information when documenting large regional floods, although in many cases the river stages and discharges are unknown (McGlashan and Briggs, 1939). In some cases, electronic databases of historical flood estimates are available, such as in Colorado, by Kohn and others (2013).

The U.S. Geological Survey Water Resources Data Reports, ~~that which~~ have been published for each state ~~(1962-2005)~~ State (1962-2005), contain some limited historical flood descriptions and information that can be extremely valuable for frequency analysis. The information is provided on the site information sheet for individual gaging stations. Since 2006, this same information can be obtained for each individual gage, if the gaging station is currently in operation. Three types of data are typically presented in the reports and site information summaries: (1) dates, stages, and sometimes discharges of observed floods prior to the gaging station period of record, ~~e.g. for example~~, Durlin and Schaffstall (2002, p. 210); (2) a large flood during the period of record that is known to be the ~~maximum stage and discharge since at least~~ maximum stage and discharge since at least some historic date, ~~e.g. for example~~, Crowfoot and others (1997, p. 413); and (3) a large flood during the period of record that is known to be the ~~maximum stage and discharge since~~ maximum stage and discharge since some historic date, ~~e.g. for example~~, May and others (1996, p. 193). The information provided in (2) and (3) sometimes only refers to either stage or discharge, depending on the observation or estimate made. In addition, there is a very subtle difference between the information provided in (2) and (3). Data provided as (2) indicate one ~~does not~~ does not have information on any flood discharges or stages prior to the date stated. One does have knowledge of a flood in the historical year stated in (3). The information for cases (1) and (3) is typically stored in electronic format in the ~~U.S. Geological Survey~~ USGS NWIS database. The data are generally summarized in the following two columns: discharge codes, where a “7” indicates that the discharge is a historic peak, and a “highest since” column, where the historic year is listed. These data need to be evaluated on an individual basis to properly estimate  $n_h$  and  $T_h$ .

State reports and publications are another major source of historical flood information. These publications can contain information on record floods, stages, historical periods, and impacts to infrastructure. For example, Suttie (1928) states “there are three great storms affecting Connecticut that are worthy of particular mention: 1869, 1897, and 1927<sup>22,23</sup>,” this information suggests that rainfall amounts and flood discharges are less than values estimated in the intervening time between these three events. For the 1869 flood, Suttie (1928, p. 120)

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

states “the Connecticut River gage at Hartford registered 26.3 feet. This is the highest stage in over a century caused by rain alone”; this information can be utilized to estimate the historical period  $h$  one may use for the 1869 stage. Many other ~~state~~ State reports contain relevant examples such as this one.

Journals and other Federal Agency reports are invaluable sources for historical flood information. The primary historical journal references are the Journal of the Boston Society of Civil Engineers and Transactions of the American Society of Civil Engineers. For example, Kuichling (~~1917, pp. 650-663~~) Kuichling (1917, p. 650-663) provides a table of maximum unit discharges for large floods in the United States to at least 1786; he also includes a reference list that includes many journals, and Geological Survey and State reports. The U.S. Army Corps of Engineers retains flood files at District offices. Community flood information and experiences are usually included in Federal Emergency Management Agency (FEMA) Flood Insurance Studies. A detailed example of historical flood data collection is provided in Thomson and others (1964); they present a flood chronology in New England from 1620 through 1955.

## Paleoflood Data and Botanical Information

Paleoflood data and botanical information for river basins and specific locations can be obtained from existing, previously ~~published~~ published sources and institutions that have obtained the data, or by field data collection at the site of interest. The main sources of existing, previously ~~published~~ published paleoflood and botanical data are various institutions that have collected the data, such as Federal agencies, ~~state~~ State agencies, and academic institutions. These data are routinely documented in journal articles, technical reports and databases, books, and some electronic databases. Over the past 20 years, the University of Arizona, ~~Bureau of~~ Bureau of Reclamation, U.S. Geological Survey, and other agencies and institutions have embarked on numerous field campaigns to obtain paleoflood data relevant for flood frequency. Similar to historical information, paleoflood and botanical data are obtained by initially searching for relevant documents and contacting institutions that have interests and facilities within the watershed of interest. After a literature search, paleoflood and botanical data can be obtained within the watershed by conducting comprehensive field studies and data collection efforts.

Several journal articles and books are key references in obtaining previously ~~published~~ published paleoflood and botanical information, and are indispensable guides for data collection efforts. Wohl and Enzel (1995) provide a useful introduction and overview of available paleoflood data. Baker and others (1988) and House and others (2002a) are key references that describe data for numerous case studies and locations, present methods for paleoflood data collection, and contains numerous citations to other relevant works and data. ~~Baker (2008, Table 2)~~ Baker (2008, table 2) summarizes many paleoflood studies and data collection completed in the United States. Benito and O'Connor (2013) and Baker (2013) provide current summaries on paleoflood and paleohydrology data and methods. Paleoflood data are readily used with EMA; for example England and others (2003a) summarized paleoflood data and demonstrated its use in flood frequency with EMA for a number of sites in the United States.

Paleoflood data for many locations within the ~~western~~ Western United States have been collected by ~~the Bureau of~~ Bureau of Reclamation for dam safety analyses. These data are typically available for many rivers and locations adjacent to Reclamation dams and other Department of Interior facilities, in order to document the most extreme floods and ~~non-exceedance~~ nonexceedance information in the Holocene. Reclamation staff typically collect paleoflood data at one of three levels: reconnaissance, intermediate, or detailed. As the level of study increases, more stratigraphic and soil-age data are obtained, and hydraulic models used to estimate discharge increase in complexity. These data are available in numerous Reclamation reports for specific projects and ~~for~~ or watersheds, and in databases (Klinger and Godaire, 2002). Some representative studies include the following: the American River and adjacent basins near Sacramento, California (Bureau of Reclamation, 2002); the North Platte River near Rawlins and Glendo, Wyoming (Levish and others, ~~2003~~ 2003); the Arkansas River near Pueblo,

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

Colorado (England and others, 2006); the South Fork Boise River, Idaho (Klinger and Bauer, 2010); the North Fork Red River near Altus, Oklahoma (Godaire and Bauer, 2012); the San Joaquin River near Fresno, California (Godaire and others, 2012), and the Rio Chama near El Vado Dam, New Mexico (Godaire and Bauer, 2013). Peak-flow frequency estimates have been made at these sites using EMA. The ~~U.S. Geological Survey~~ [USGS](#) has also conducted numerous paleoflood studies using reconnaissance or regional approaches (Jarrett and Tomlinson, 2000) and detailed methods for flood hazard assessments at specific locations (Harden and others, 2011). Some paleoflood data are available in electronic databases, such as Kohn and others (2013).

Botanical information and data, such as tree scars and tree rings, are available in publications and some electronic databases. Some essential publications on methods and data are Hupp (1987, 1988), and Yanosky and Jarrett (2002); these contain numerous citations to other relevant works and data. McCord (1990) provides tree-scar data at select sites in Arizona, Utah, New Mexico, and Colorado. Additional resources include the Laboratory of Tree-Ring Research at the University of Arizona and the International Tree-Ring Data Bank.

## Regional Information

Regional information that can be considered for flood frequency typically consists of regional estimates of flow statistics. Regional skew coefficient  $G$  estimates and mean-square error  $MSE_G$  estimates can be obtained for many locations in current ~~U.S. Geological Survey~~ [USGS](#) flood frequency reports for regions or individual ~~states~~ [States](#). For example, regional skew estimates are available for the Southeastern ~~US~~ [United States](#) (Gotvald and others, 2009; Feaster and others, 2009; Weaver and others, 2009), California (Parrett and others, 2011; Gotvald and others, 2012), Iowa (Eash and others, 2013), Arizona (Paretti and others, 2014b), Missouri (Southard and Veilleux, 2014), and Vermont (Olson, 2014). The USGS is in the process of updating regional skew estimates for many other ~~states~~ [States](#). Regional flood quantile estimates  $Q_i$  and their variances  $V_{reg,i}$  are also available in these reports, and are useful in record extension (~~Appendix~~ [appendix](#) 8) and in weighting of independent estimates (~~Appendix~~ [appendix](#) 9). These flood frequency reports and additional information on regional skew and regional quantile estimates for many locations are available from the ~~U.S. Geological Survey~~ [USGS](#) and the HFAWG at <https://acwi.gov/hydrology/Frequency/b17c/>.

In lieu of current published estimates, it is recommended that users consult with the ~~U.S. Geological Survey~~ [USGS](#) to determine the availability of regional skew estimates that have been prepared using current methods, described in the ~~Section~~ [section](#) Estimating Regional Skew. The regional skew estimates published in ~~IACWD (1982, Plate 1)~~ [IACWD \(1982, plate 1\)](#) are not recommended for use in flood frequency studies. When no other regional skew information is available, it is recommended that new estimates be developed for the region of interest.

## Precipitation and Climate Information

Precipitation information that is potentially useful for flood rainfall-runoff modeling and flood frequency analysis is generally available from various Federal and state agencies. Point precipitation data and radar rainfall products are available from the National Oceanic and Atmospheric Administration (NOAA) ~~;~~ [and](#) National Centers for Environmental Information (NCEI). National Weather Service River Forecast Centers also provide multisensor precipitation (combined radar and precipitation gage) estimates (MPE) across the United States. Precipitation frequency estimates and time series are available from the National Weather Service Hydrometeorological Design Studies Center. Precipitation data for many of the largest historical rainfall events and floods can be obtained from extreme storm catalogs at the U.S. Army Corps of Engineers, ~~Bureau of~~ Reclamation, and through the Extreme Storm Events Working Group at <https://acwi.gov/hydrology/extreme-storm/index.html>. Precipitation and temperature data important for rainfall-runoff modeling of extreme floods can be obtained

from the National Resources Conservation Service (NRCS) Snow Telemetry (SNOTEL) and snow course data. The National Weather Service's National Operational Hydrologic Remote Sensing Center (NOHRSC) SNOw Data Assimilation System (SNODAS), available through the National Snow and Ice Data Center, is another valuable [data-set-dataset](#) for snow cover and associated variables.

Climate information that is useful for a hydroclimatological perspective on floods is available from the NOAA Earth System Research Laboratory (ESRL); other sources may be found through NOAA and at <https://acwi.gov/hydrology/Frequency/b17c/>. Information on climate models, downscaling information, and climate change, that is potentially relevant for floods, is under rapid development and has not been comprehensively evaluated for use in flood frequency studies. An overview is presented in Brekke and others (2009). Downscaled climate information and tools for climate change assessment studies are available from various sources, such as the Bureau of Reclamation at <http://www.usbr.gov/climate>, the USGS at <https://cida.usgs.gov/gdp/>, and the USACE at <http://www.corpsclimate.us/index.cfm>. Additional resources may be found at <https://acwi.gov/hydrology/Frequency/b17c/>.

## Data Representation

As described in the [Section-section](#) Data Representation using Flow Intervals and Perception Thresholds, a generalized representation of peak-flow data are used in flood frequency analysis ~~to~~ to describe what is known about annual peak flows in a given year  $Y$ , over a range of years  $n$ . Typical peak-flow intervals and thresholds for systematic (gage) records, crest-stage gages, historical data, and PILFs are summarized in [Table-table](#) 3.1.

At gaging stations (systematic record site), typical data consist of flow intervals and perception intervals that are shown in [Figure-figure](#) 3.1. Flows each year are typically known [as](#) and are considered “point” estimates. Because there is a gaging station and continuous observations, one can observe or measure a flow of any magnitude. Perception thresholds are typically 0 to  $\infty$ . On occasion, there may be flows best described with intervals or binomial observations, where one knows flows exceeded some lower estimate  $Q_{Y,lower} = Q_Y$ . In the case of a broken record, where the gage is discontinued or temporarily stopped for a period, one has no information or knowledge. For such broken record years, flows and perception thresholds are set equal to  $\infty$ , unless one has additional information from a historical period or large flood, as shown in [Figure-figure](#) 9. A systematic record example is shown in [Figure-10.1 in Appendix-figure 10.1 in appendix 10](#).

The flow intervals and perception thresholds for a crest-stage gage are shown in [Figure-figure](#) 3.2. In this situation, flows are recorded only if they exceed a [gage-base-gage-base](#) discharge  $Q_b$ . This is represented by a perception threshold  $T_b$ . Perception thresholds are typically  $T_b$  to  $\infty$ . On occasion, there may be flows best described with intervals or binomial observations, where one knows flows exceeded some lower estimate  $Q_{Y,lower} = Q_Y$ . A crest-stage gage example is shown in [Figure-10.10 in Appendix-figure 10.10 in appendix 10](#).

The typical flow intervals and perception thresholds for historical data are shown in [Figure-figure](#) 3.3. In this situation, flows are recorded only if they exceed a historical perception threshold  $T_h$ . There may be more than one perception threshold to describe historical and ~~for(or)~~ paleoflood periods, as shown in [Figure-figure](#) 12 in the [Section-section](#) Data Representation using Flow Intervals and Perception Thresholds. Perception thresholds are typically  $T_h$  to  $\infty$ . Upper perception thresholds may be less than infinity where high-water marks from very high stages may not be preserved, especially in canyon environments. Within the historical period, there may be extreme floods that are best described with intervals or binomial observations, where one knows flows exceeded some lower estimate  $Q_{Y,lower} = Q_Y$ . Historical data examples are shown in [Figures-figures](#) 10.8 and 10.12, and with historical and paleoflood data shown in [Figure-10.14 in Appendix-figure 10.14 in appendix 10](#). It is important to estimate the historical period(s) and perception threshold(s), as described in the [Section-section](#) Data Representation using Flow Intervals and Perception Thresholds.

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

**Table 3.1.** Generalized data representation of peak-flow intervals and perception thresholds for each year - (typical values).

<del>Data Source</del> <u>source</u>	<del>Data Type</del> <u>type</u>	<del>Flow Interval</del> <u>interval</u>	<del>Perception Threshold</del> <u>threshold</u>
<del>gage</del> <u>Gage</u>	<del>point</del> <u>Point</u>	$Q_{Y,lower} = Q_Y$ $Q_{Y,upper} = Q_Y$	$T_{Y,lower} = 0$ $T_{Y,upper} = \infty$
<del>gage</del> <u>Gage</u>	<del>interval</del> <u>Interval</u>	$Q_{Y,lower}$ $Q_{Y,upper}$	$T_{Y,lower} = 0$ $T_{Y,upper} = \infty$
<del>gage</del> <u>Gage</u>	<del>binomial</del> <u>Binomial</u>	$Q_{Y,lower}$ $Q_{Y,upper} = \infty$	$T_{Y,lower} = 0$ $T_{Y,upper} = \infty$
<del>gage</del> <u>Gage</u> ~ record	<del>broken</del> <u>Broken</u>	$Q_{Y,lower} = \infty$ $Q_{Y,upper} = \infty$	$T_{Y,lower} = \infty$ $T_{Y,upper} = \infty$
<del>crest</del> <u>Crest</u> ~ stage	<del>point</del> <u>Point</u>	$Q_{Y,lower} = Q_Y$ $Q_{Y,upper} = Q_Y$	$T_{Y,lower} = Q_b$ $T_{Y,upper} = \infty$
<del>crest</del> <u>Crest</u> ~ stage	<del>interval</del> <u>Interval</u>	$Q_{Y,lower} > Q_b$ $Q_{Y,upper}$	$T_{Y,lower} = Q_b$ $T_{Y,upper} = \infty$
<del>historical</del> <u>Historical</u>	<del>point</del> <u>Point</u>	$Q_{Y,lower} = Q_Y$ $Q_{Y,upper} = Q_Y$	$T_{Y,lower} = Q_h$ $T_{Y,upper} = \infty$
<del>historical</del> <u>Historical</u>	<del>interval</del> <u>Interval</u>	$Q_{Y,lower} > Q_h$ $Q_{Y,upper}$	$T_{Y,lower} = Q_h$ $T_{Y,upper} = \infty$
<del>historical</del> <u>Historical</u>	<del>binomial</del> <u>Binomial</u>	$Q_{Y,lower} > Q_h$ $Q_{Y,upper} = \infty$	$T_{Y,lower} = Q_h$ $T_{Y,upper} = \infty$
<del>historical</del> <u>Historical</u>	<del>censored</del> <u>Censored</u>	$Q_{Y,lower} = 0$ $Q_{Y,upper} = Q_h$	$T_{Y,lower} = Q_h$ $T_{Y,upper} = \infty$
PILF	<del>censored</del> <u>Censored</u>	$Q_{Y,lower} = 0$ $Q_{Y,upper} = Q_l$	$T_{Y,lower} = Q_l$ $T_{Y,upper} = \infty$

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017



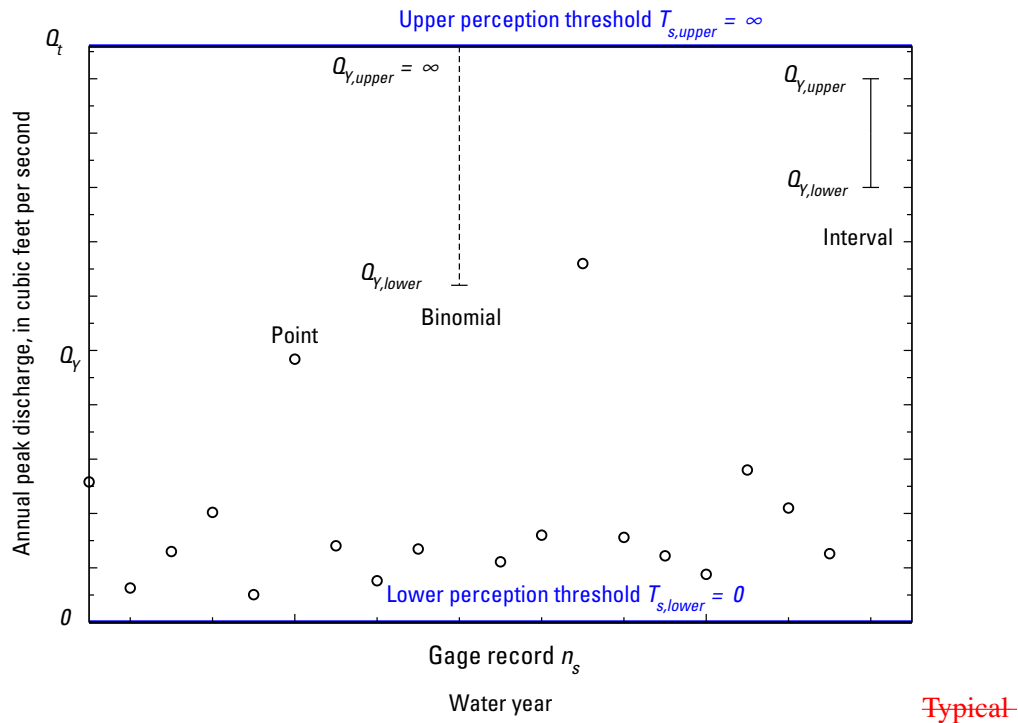


Figure 3.1. Graph showing typical flow intervals and perception thresholds for a gaging station.

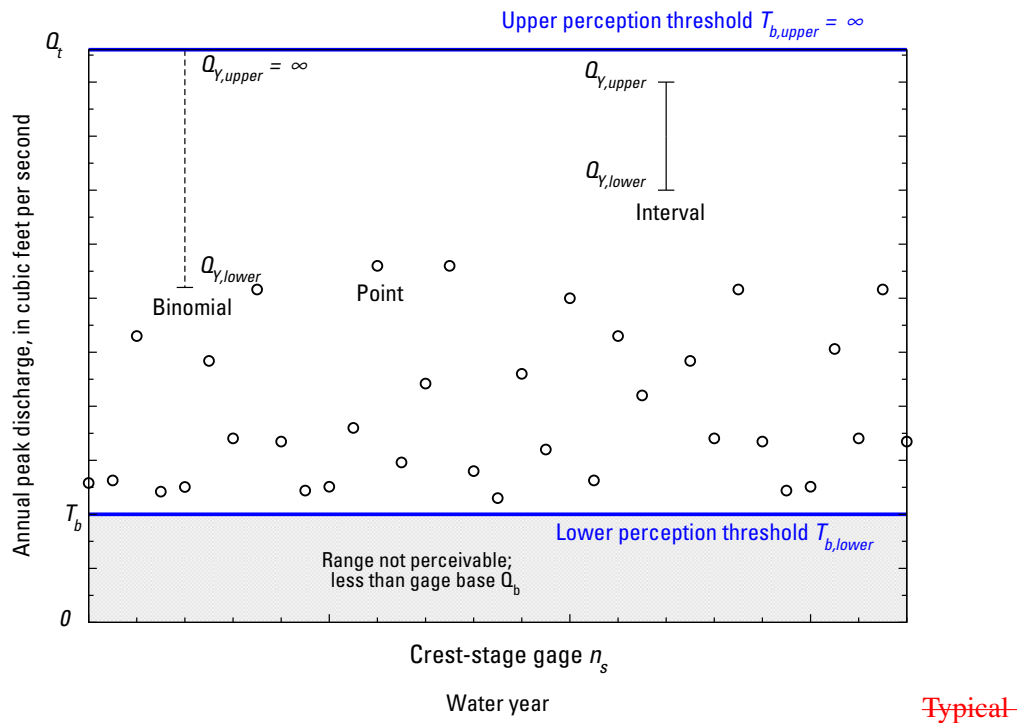
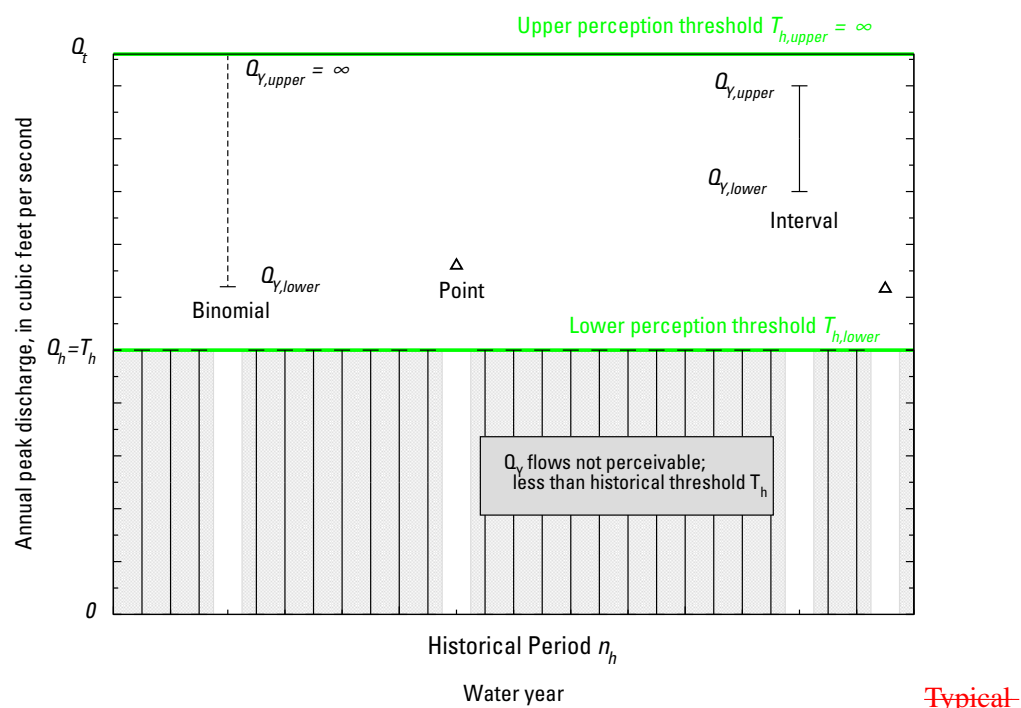


Figure 3.2. Graph showing typical flow intervals and perception thresholds for a crest-stage gage.



**Figure 3.3.** Graph showing typical flow intervals and perception thresholds for historical data.

DRAFT: August 31, 2017

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

## Appendix 4. Initial Data Analysis

---

1

2

DRAFT: August 31, 2017

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

## Initial Data Analysis

When conducting a flood frequency analysis, a first step is to undertake basic analysis of the peak-flow time series to check for obvious errors and to check that the data conform to the assumptions used in the frequency analysis. One of the main assumptions in flood frequency analysis is that the data in the peak-flow time series are independent and identically distributed. Some tests that check that these assumptions are reasonable for a particular time series include tests for autocorrelation and nonstationarity. Visual inspection of the time series can also reveal issues that need to be addressed.

## Visual Inspection: Plot the data

Before any formal statistical tests are employed, a visual inspection of a plot of the peak-flow time series can be used to help identify any potential errors with the data. For example, any peaks that are orders of magnitude different from the others should be verified. Visual inspection of the time-series plot may also reveal obvious changes in the mean or standard deviation of the peak-flow data over time. For example, construction of a dam and reservoir may drastically alter peak-flow time series and the entire pre- and post-dam time series should not be used together for a peak-flow frequency analysis.

## Autocorrelation

It is recommended that an annual flood series be examined for autocorrelation through the use of a correlogram (Salas, 1993). In an autocorrelated time series, the value in one time step is correlated with the value in a previous (and future) time step. Autocorrelated time series can also be said to exhibit persistence. Hydrologic time series will often exhibit long-term persistence. Note that this can affect trend testing, as discussed below.

## Trends and shifts

The peak-flow frequency analysis methods described in this document are only applicable when the peak-flow data are believed to be part of the same underlying population. Changes in peak-flow generation processes can lead to gradual trends or abrupt shifts in the peak-flow time series. Statistical tests for trends and shifts can be useful for detecting such changes in the peak-flow time series. Depending on the likely causes and the magnitude of any detected changes, different treatments may be needed before Bulletin 17C methods can be applied. These will be discussed in a future update to this document. A particularly difficult case is when it is unknown whether the apparent trend will continue, level off, or reverse in the future. Possible approaches for dealing with such changes have been discussed in the research literature, but a consensus on best practices has not yet emerged. Consequently, substantial judgment must be exercised when trends are found.

Changes may occur gradually or abruptly and different tests are commonly used to test for the presence of either type of change. A visual inspection of a plot of the annual peak-flow time series should be the first step in assessing the time series. It is recommended that this be followed by trend tests to help assess whether changes over time may be important for the flood frequency analysis. This can be followed by a change point test for an abrupt change, if desired. The specific tests described below have been used frequently, but other tests may also be considered.

Trend tests and change point analysis are most commonly done on the mean values of a time series, but tests for change in the variance of a time series can also be considered. Note that these tests can be sensitive to the start and end points used in the analysis. For example, if the period of record happens to either start or end with a large peak, there may be an apparent trend in the data. However, this apparent trend may simply be the result

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017



of the particular sample that was used. A slightly longer or shorter record would not show the same apparent trend. In other cases, the period of record may include only the drying or wetting phase of an oscillation with long periodicity. The apparent trend results from a finite record length, but in this case a much longer period of record is needed to fully understand the data.

## Statistical tests

A common test for trends in a time series is the Mann-Kendall test. This test uses Kendall's  $\tau$  as the test statistic to measure the strength of the monotonic relationship between annual peak streamflow and the year in which it occurred. The Mann-Kendall test is nonparametric and does not require that the data conform to any specific statistical distribution. The statistic is calculated using the difference between concordant and discordant data pairs of the observed streamflow peaks with time, and not the actual magnitudes of the data values (Helsel and Hirsch, 1992). Positive values for  $\tau$  indicate that occurrences of annual peak streamflows are increasing with time for the period of record, while negative values of  $\tau$  indicate that annual peak streamflows are decreasing with time for the period of record.

As with other statistical tests, a  $p$ -value can be calculated for the test. Note that the  $p$ -values will be correct only when there is no serial correlation in the annual time series. This requirement can be problematic for hydrologic time series, which can exhibit short-term and long-term persistence (Cohn and Lins, 2005).

In addition to the statistical significance of a trend, the actual magnitude of the trend should be considered. The Theil slope (Helsel and Hirsch, 1992, p. 266) can be calculated in conjunction with Kendall's  $\tau$  for this purpose. It is calculated as the median of all the slopes calculated by using all the possible pairs of peak-flow values and years, and is a nonparametric estimate of the slope.

In some situations, there may be an abrupt shift (McCabe and Wolock, 2002) or change in the time series, rather than a gradual trend. For example, there may be distinct periods, exhibiting different flood characteristics, before and after installation of flood-control structures. In other cases, the reason for the step change may not be as evident, but abrupt changes may still be found. Villarini and others (2009a), for example, found step changes that appeared to coincide with changes in the streamgauge location. To refine the analysis, the test for a monotonic trend could be followed by a test for a step change. The Wilcoxon rank-sum test (also known as the Mann-Whitney test) or the Kolmogorov-Smirnov test are both nonparametric tests that can be used to test for differences between two samples, when there is a suitable hypothesis for separating the time series into two or more sections (Helsel and Hirsch, 1992). The potential step change should not be identified solely on the basis of visual inspection of the data, as this biases the test. The Pettitt test (Pettitt, 1979; Mallakpour and Villarini, 2016) and Lombard's Smooth Change Model (Lombard, 1987; Nayak and Villarini, 2016) have both been suggested as alternative tests for abrupt changes that do not require an analyst to predetermine where a likely change occurs (Villarini and others, 2009a; Quessy and others, 2011). Additional information on these tests can be found in Helsel and Hirsch (1992) and other statistical textbooks. Some information on changepoint tests is available in supplementary material on the Bulletin 17C website at <https://acwi.gov/hydrology/Frequency/b17c/>.

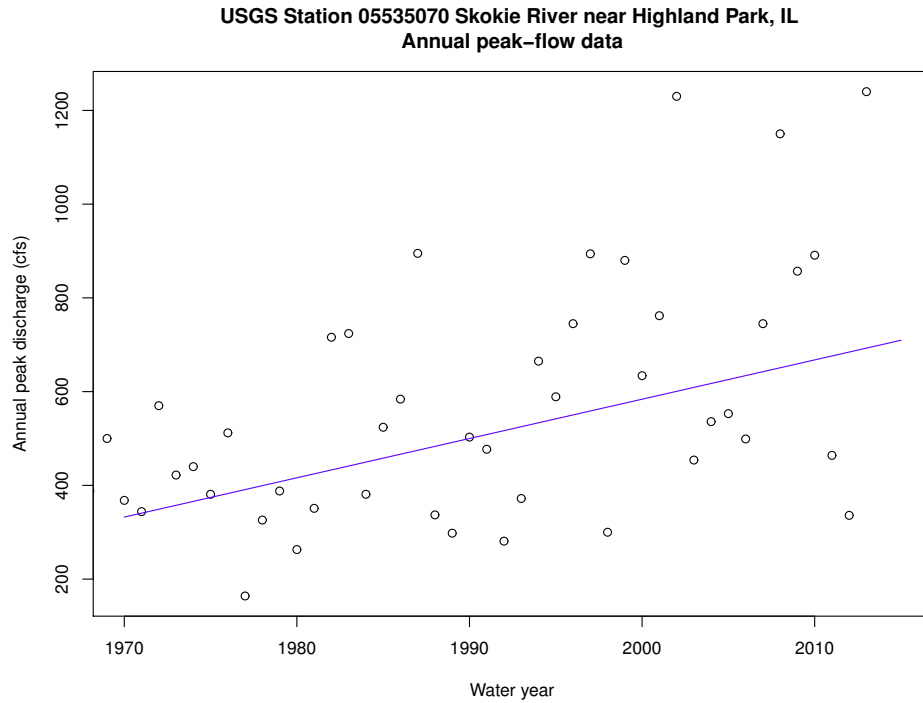
## Example: Skokie River near Highland Park, IL

This example uses data from USGS gaging station 05535070, Skokie River near Highland Park, Illinois. Figure 4.1 shows a time-series plot of the Skokie River. It is a 21.1-square-mile watershed that has become more and more urbanized over time. The urbanized fraction was about 0.60 at the beginning of the period of record in 1967, and increased to about 0.90 by 2014 (Over and Soong, 2015). Visual inspection of the time series reveals an increasing trend over time.

The visual trend is confirmed with the Mann-Kendall test. The results from the test are as follows:

PROVISIONAL  
THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.  
IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).  
IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017



**Figure 4.1.** Graph showing U.S. Geological Survey station 05535070 Skokie River near Highland Park, Illinois (IL) time-series plot. Annual peak discharges have increased at this streamgage due to urbanization. The line is the fitted Theil line with slope 8.4 cubic feet per second (cfs) per year.

$$\tau = 0.321,$$

$$p\text{-value} = 0.00156, \text{ and Theil slope} = 8.4 \text{ cfs/year.}$$

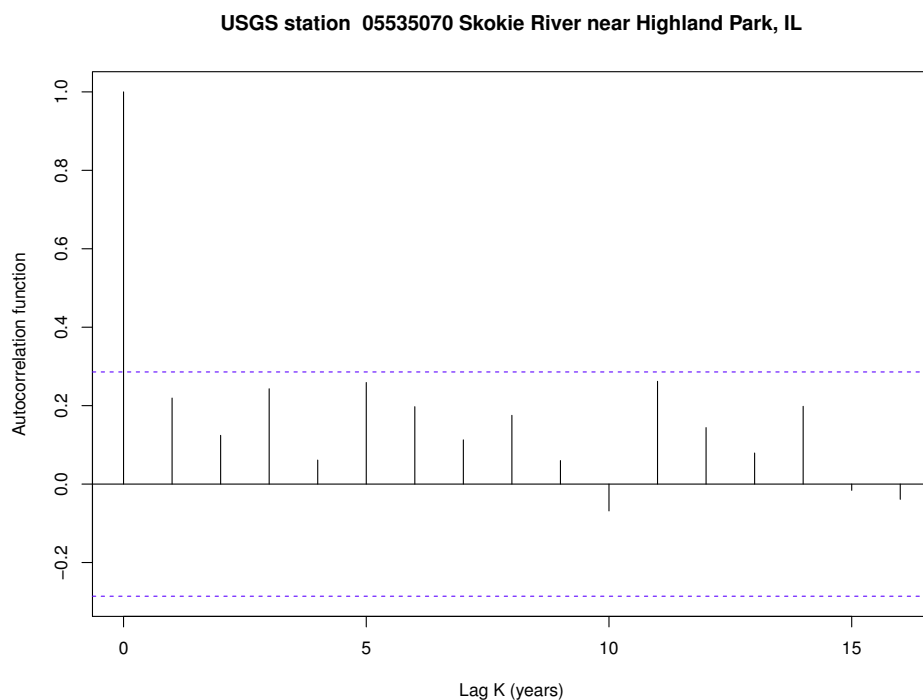
The annual peak flows at this station are not significantly autocorrelated, as shown in figure 4.2. This indicates that the estimated  $p$ -value is appropriate and is unaffected by autocorrelation.

1  
2  
3  
4

**DRAFT: August 31, 2017**

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.  
IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).  
IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.



**Figure 4.2.** Graph showing U.S. Geological Survey station 05535070 Skokie River near Highland Park, Illinois (IL) autocorrelation plot. The annual peaks do not exhibit any statistically significant autocorrelation for lag times between 1 and 15 years. The dashed lines are the thresholds for significant autocorrelation.

DRAFT: August 31, 2017

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.  
IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).  
IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

## Appendix 5. Threshold-Exceedance Plotting Positions

---

1

2

DRAFT: August 31, 2017

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

## Threshold-Exceedance Plotting Positions

This appendix provides an overview and equations for threshold-exceedance-based-exceedance-based plotting positions. Table 5.1 provides plotting position parameters  $a$  and their motivation. A plotting parameter  $a = 0.0$ , corresponding to a Weibull formula, is recommended as a default value, consistent with current practice. Other plotting parameters, including 0.40 (Cunnane), 0.44 (Gringorten), and 0.50 (Hazen), are traditional choices that may also be considered. Some examples are shown in Appendix appendix 10.

Consider a historical flood record with an  $n_h$ -year historical period in addition to a complete  $n_s$ -year gaged flood record. Assume that during the total  $n = (n_s + n_h)$  years of record, a total of  $k$  floods exceeded a perception threshold for historical floods (Figure fig. 3). If the  $k$  values which that exceeded the threshold are indexed by  $i = 1, \dots, k$ , reasonable plotting positions approximating the exceedance probabilities with the interval  $(0, p_e)$  are

$$p_i = p_e \left( \frac{1-a}{k+1-2a} \right) = \frac{k}{n} \left( \frac{i-a}{k+1-2a} \right) \quad (5.1)$$

where  $a$  is a value from Table table 5.1 and  $p_e = k/n$  is the probability of exceeding a threshold. For  $k \gg (1-2a)$ ,  $p_i$  is indistinguishable from  $\frac{i-a}{n+1-2a}$  for a single threshold. Hirsch (1987) notes that for the first  $k$  floods, equation (5.1) 5.1 is identical to the Hazen formula with  $a = 0.5$ , and is very close to the Gringorten formula with  $a = 0.44$ . Reasonable choices for  $a$  generally make little difference to the resulting plotting positions.

The plotting positions for systematic record floods below the threshold must be adjusted to reflect the additional information provided by the historical flood record, if the historical flood data and the systematic record are to be analyzed jointly in a consistent and statistically efficient manner (Hirsch and Stedinger, 1987). In this case, let  $e_s$  be the number of gaged-record floods that exceeded the threshold and hence, hence, are counted among the  $k$  exceedances of that threshold. Plotting positions within  $(p_e, 1)$  for the remaining  $(n_s - e_s)$  below-threshold gaged-record floods are

$$p_r = p_e + (1 - p_e) \left( \frac{r-a}{n_s - e_s + 1 - 2a} \right) \quad (5.2)$$

for  $r = 1$  through  $n_s - e_s$ , where again  $a$  is again a value from Table table 5.1.

This approach directly generalizes to several thresholds. For the multiple exceedance threshold cases shown in Figure figure 12, equation (5.1) 5.1 can be generalized (Hirsch and Stedinger, 1987; Stedinger and others, 1988, 1993). The number of thresholds is defined as  $j$  ( $j = 1, \dots, m$ ), where the thresholds  $Q_j$  ( $j = 1, \dots, m$ ) are ordered (sorted) from largest to smallest such that  $Q_1 > Q_2 > \dots > Q_m$ . The probability of exceedance  $p_{e_j}$  for each threshold  $j$  is defined as ÷

$$p_{e_j} = p_{e_{j-1}} + (1 - p_{e_{j-1}}) q_{e_j} \quad (5.3)$$

where  $q_{e_j}$  is the conditional probability that a flood falls between the  $j^{th}$  and  $(j-1)^{th}$  threshold. It is defined by ÷

$$q_{e_j} = \frac{k_j}{n_j - \sum_{l=1}^{j-1} k_l} \quad (5.4)$$

where  $k_j$  is the number of floods that exceed threshold  $j$  but not any higher thresholds  $(j-1)$ , and the denomi-



**Table 5.1.** Typical plotting position parameter  $a$  values and their motivation (Stedinger and others, 1993).

Method	$a$	Motivation
Weibull	0.00	Unbiased exceedance probabilities for all distributions
Cunnane	0.40	Approximately quantile-umbiased
Gringorten	0.44	Optimized for Gumbel distribution
Hazen	0.50	A traditional choice

1 nator in equation ~~(5.4)~~ 5.4 is the number of years ( $n_j$ ) that threshold  $Q_j$  applies minus the sum of all floods  $k_l$   
2 that exceed any higher ( $j-1, j-2, \dots$ ) thresholds during period  $n_j$ . The above-threshold floods may be plotted  
3 by  $\div$

$$4 \quad p_i = p_{e_{j-1}} + (1 - p_{e_{j-1}}) q_{e_j} \left( \frac{i - a}{k_j + 1 - 2a} \right) \quad (5.5)$$

5 and the floods below all thresholds ( $k_j + 1, \dots, g$ ) can be plotted using equation ~~(5.2)~~ 5.2 with  $p_e$  equal to  $p_{e_j}$ .

## Appendix 6. Potentially Influential Low Floods

---

1

2

DRAFT: August 31, 2017

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

## Potentially Influential Low Floods

This appendix provides a general introduction to Potentially ~~Influential~~ Influential Low Floods (PILFs), and describes computation details for identifying PILFs — the Multiple Grubbs-Beck test (MGBT). Some details on MGBT and its performance are documented in Cohn and others (2013), Lamontagne and others (2013), and Lamontagne and others (2016). Some examples of detecting PILFs with the MGBT are provided in [Appendix appendix 10](#).

## PILF Background and Philosophy

There are recognized problems with the Grubbs-Beck (GB) test (Grubbs and Beck, 1972) used in Bulletin 17B (IACWD, 1982) when there are multiple low outliers. This issue was discussed in the Bulletin 17B FAQ, under the section “Low Outliers” written by Bill Kirby. Relevant portions of that FAQ are reproduced here, with *emphasis emphasis* on the important issues.

~~Bulletin-17-B Bulletin 17B~~ detects low outliers by means of a statistical criterion (the ~~Grubbs-Beck-GB~~ test) rather than by consideration of the influence of low-lying data points on the fit of the frequency curve. The test is based on the standardized distances,  $(x_i - \hat{\mu})/\hat{\sigma}$ , between the lowest observations and the mean of the ~~data set. The test is easily defeated by occurrence of multiple low outliers, which exert a large distorting influence on the fitted frequency curve, but also increase the standard deviation,  $\hat{\sigma}$ , thereby making the standardized distance too small to trigger the Grubbs-Beck test.~~ dataset. The test is easily defeated by occurrence of multiple low outliers, which exert a large distorting influence on the fitted frequency curve, but also increase the standard deviation,  $\hat{\sigma}$ , thereby making the standardized distance too small to trigger the Grubbs-Beck test.

The FAQ also provides further background, and a hydrological basis to deviate from the GB test as follows, with *emphasis emphasis* on the relevant text.

Obviously, the intention is to allow as many low outliers to be designated as necessary to achieve a good fit to the part of the ~~data-set-dataset~~ that contains the significant flood and near-flood events. Equally obviously, the intention is that the ~~Grubbs-Beck-GB~~ result be used unless the resulting poor fit gives compelling justification for not doing so. There is no universal method that can be followed blindly to achieve a good fit. The sensitivity analysis alluded to in Bulletin ~~17-B is based on the engineering-hydrologic-common-sense proposition that the smallest observations in the data set do not convey meaningful or valid information about the magnitude of significant flooding, although they do convey valid information about the frequency of significant flooding. Therefore, if the upper tail of the frequency curve is sensitive to the numerical values of the smallest observations, then that sensitivity is a spurious artifact based on the mathematical form of the assumed but in fact unknown flood distribution, and has no hydrologic validity.~~ 17B is based on the engineering-hydrologic-common-sense proposition that the smallest observations in the dataset do not convey meaningful or valid information about the magnitude of significant flooding, although they do convey valid information about the frequency of significant flooding. Therefore, if the upper tail of the frequency curve is sensitive to the numerical values of the smallest observations, then that sensitivity is a spurious artifact based on the mathematical form of the assumed, but in fact unknown, flood distribution, and has no hydrologic validity.

Others have noted this hydrologic phenomenon. A key observation is from Klemeš (1986, p. 183S), reproduced as follows: “For it is by no means hydrologically obvious why the regime of the highest floods should be affected by the regime of flows in years when no floods occur, why the probability of a severe storm hitting this basin should depend on the accumulation of snow in the few driest winters, why the return period of a given heavy rain should be by an order of magnitude different depending, say, on slight temperature fluctuations during the melting seasons of a couple of years (p. 183S).”

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

Klemeš (2000, p. 229) also described this hydrological problem in the context of frequency distributions, as follows, with *emphasis* on the relevant text.

“... It is ironic that the only clue the FA (Frequency Analysis) theory inadvertently takes from hydrology is the wrong one. It derives the “distributional assumptions” [i.e., the general shape of  $F(X)$ ] from a “probability plot” such as Fig. 1(b) whose shape is dominated by the small and medium observations. This shape is generally convex on the Gaussian plot, because hydrological phenomena like precipitation, runoff, snow cover, etc., have a zero lower bound, which “bends” the lower tail of the plot towards a horizontal asymptote. As a result, all the “standard” distribution models are convex on Gaussian frequency scale; they all are models with positive skewness. *Hence, it is the physical regime prevailing in the formation of the lower tail that determines the shape of the extrapolated upper tail; observations that are hydrologically least relevant to the high extremes and to the safety of facilities affected by them — have the greatest influence on their estimated “probabilities”! ...*”

These observations, as well as the data issues described in the [Section-section](#) Zero Flows and [Potentially-Influential](#) [Potentially Influential](#) Low Floods, are handled with the MGBT.

## Computational Details for Identifying PILFs with MGBT

The purpose of using the MGBT is to identify PILFs. PILFs are small observations (or zero flows) that potentially have a large influence on the fitted frequency curves. When [data-sets-datasets](#) are negatively skewed, the smallest observations can be very influential in determining the estimated skewness coefficient and the estimated 1% AEP flood. The new MGBT is a statistically appropriate generalization of the GB test that is sensitive to the possibility that several of the smallest observations are “unusual,” or are potentially very influential. The MGBT also correctly evaluates cases where one or more observations are zero, or are below a recording threshold (partial record sites). Thus, it provides a consistent, objective, and statistically defensible algorithm that considers whether a range of the smallest observations should be classified as PILFs for a wide range of situations that are observed in practice. See, for example, cases in Lamontagne and others (2012), Paretti and others (2014a), and examples in [Appendix-appendix](#) 10.

To provide an objective criteria for multiple low outlier identification, MGBT employs the actual distribution of the  $k^{th}$  largest observation in a sample of  $n$  independent normal variates, where the probability  $p_{[k:n]}$  that the  $k^{th}$  largest observation in a normal sample of size  $n$  might have appeared to be smaller than the value observed (Cohn and others, 2013). If  $p_{[k:n]}$  is small, then the  $k^{th}$  observation is unusually small.

To test  $H_0$ , we consider whether  $\{X_{[1:n]}, X_{[2:n]}, \dots, X_{[n:n]}\}$  are consistent with a normal distribution and the other observations in the sample by examining the statistic

$$\tilde{\omega} \equiv \frac{X_{[k:n]} - \hat{\mu}_k}{\hat{\sigma}_k} \quad (6.1)$$

where  $X_{[k:n]}$  denotes the  $k$ -th smallest order statistic in the sample, and

$$\hat{\mu}_k = \frac{1}{n-k} \sum_{j=k+1}^n X_{[j:n]} \quad (6.2)$$

$$\hat{\sigma}_k^2 = \frac{1}{n-k-1} \sum_{j=k+1}^n (X_{[j:n]} - \hat{\mu}_k)^2. \quad (6.3)$$

The partial mean ( $\hat{\mu}_k$ ) and partial variance ( $\hat{\sigma}_k^2$ ) are computed based on all observations larger than  $X_{[k:n]}$  to avoid swamping (specifying too many outliers). Each observation  $X_{[k:n]}$  is tested sequentially. These larger

observations greater than  $X_k$  are not suspected of being low outliers, thus  $\hat{\mu}_k$  and  $\hat{\sigma}_k^2$  are assumed to correspond to the population of interest. From  $\tilde{\omega}$ , we calculate the  $p$ -value: the probability given  $H_0$  of obtaining a value of  $\tilde{\omega}_{[k:n]}$  as small or smaller than that observed in the sample. The  $p$ -value of interest is given by

$$p_k[\eta] \equiv P[\tilde{\omega}_{[k:n]} < \eta]. \quad (6.4)$$

Substituting the definition of  $\tilde{\omega}_{[k:n]}$  from equation 6.1 and rearranging the terms yields (Cohn and others, 2013)

$$p_k[\eta] = P\left[\left(\frac{Z_{[k:n]} - \hat{\mu}_{Z,k}}{\hat{\sigma}_{Z,k}}\right) < \eta\right] \quad (6.5)$$

where  $Z_{[k:n]}$  is the  $k$ -th order statistic in a standard normal sample of size  $n$ , and  $\hat{\mu}_{Z,k}$ ,  $\hat{\sigma}_{Z,k}$  are the partial mean and standard deviation of the normal sample. If that  $p$ -value is small (for example, less than  $\alpha = 10\%$ ), then the  $k$  smallest observations are declared PILFs, such as those shown in Figure-figure 11. The PILF threshold  $X_l$  that is used in EMA is set to the  $(k+1)$ -th value.

The MGBT for identifying ~~Potentially Influential Low Floods~~ PILFs has two steps. The input data are base-10 logarithms  $X_j$  of annual peak flows from the systematic (gaging) record ( $n_s$ ), with flow values exactly known as point observations ( $Q_{Y,lower} = Q_{Y,upper} = Q_Y$ ). Flows are ranked from smallest to largest, as noted in the ~~Section~~ section Zeros and Identifying Potentially Influential Low Floods.

1. Starting at the median and sweeping ~~outward~~ outward towards the smallest observation, each observation  $X_{[k:n]}$  is tested and is identified as an outlier if  $p(k;n) \leq \alpha_{out}$ . If the  $k^{th}$  largest observation is identified as a low outlier, the outward sweep stops and the  $k^{th}$  and all smaller observations (~~i.e. that is~~, for all  $j \leq k$ ) are also identified as low outliers.
2. An ~~inward~~ inward sweep starts at the smallest observation  $X_{[1:n]}$  and moves towards the median, where the  $j^{th}$  observation is identified as an outlier if  $p(k;n) \leq \alpha_{in}$ . If an observation  $m = 1, 2, \dots, n/2$  fails to be identified as an outlier by the inward sweep, the inward sweep stops.

The number of PILFs identified by the procedure is then the larger of  $k$  and  $m - 1$ .

The algorithm has two parameters: an ~~outward-sweep~~ outward sweep significance level  $\alpha_{out}$ , and an ~~inward-sweep~~ inward sweep significance level  $\alpha_{in}$ . The recommended values used in MGBT are  $\alpha_{out} = 0.005$  (0.5%) and  $\alpha_{in} = 0.10$  (10%). These values were determined through extensive testing and evaluation by the HFAWG through careful examination of 82 sites (Cohn and others, 2017), testing and performance of alternatives (Lamontagne and others, 2013), and further investigations (Lamontagne and others, 2016).

The ~~outward~~ outward sweep seeks to determine if there is some break in the lower half of the data that would suggest the sample is best treated as if it had a number of low outliers. The ~~inward~~ inward sweep using a less severe significance level,  $p(k;n) \leq 10\%$ , mimics Bulletin 17B's willingness to identify one or more of the smallest observations as low outliers so that the analysis is more robust. Bulletin 17B also used a 10% significance test with its single GB threshold. However, a critical difference is that the MGBT ~~inward~~ sweep uses the  $p(k;n)$  function, which correctly describes whether the  $k^{th}$  largest observation in a normal sample of  $n$  variates is unusual.

For example, if a record has ~~5-five~~ zero flows, then the smallest ~~non-zero~~ nonzero flow is considered to be the ~~6<sup>th</sup>~~ sixth smallest observation in the record. This correctly reflects the fact that the flood record included ~~5-five~~ smaller values. The GB test in Bulletin 17B includes no mechanism for correcting its threshold when testing the smallest ~~non-zero~~ nonzero flood value in a record containing zeros, or below-threshold discharges at sites with crest-stage gages. This is particularly problematic because sites with zero flows are very likely to

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.



#### 94 Guidelines for Determining Flood Flow Frequency – Bulletin 17C

- 1 include one or more very small or near-zero flood values, which should legitimately be identified as low outliers
- 2 were a statistically appropriate threshold employed. The MGBT solves this problem.
- 3 Computer programs (see the ~~Section~~section Software and Examples) are used to perform the MGBT and
- 4 report critical values and PILFs.

DRAFT: August 31, 2017

---PROVISIONAL---  
THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.  
IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).  
IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

## Appendix 7. Expected Moments Algorithm (EMA)

---

1

2

DRAFT: August 31, 2017

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

## Expected Moments Algorithm (EMA)

This appendix describes features of EMA, including some computation details, a generalized expected moments algorithm, and uncertainty of EMA moments, and confidence intervals with EMA.

### EMA Computational Details

The EMA moments for the general situation with a historical flood perception threshold  $X_h$  and a PILF threshold  $X_l$  are as follows:

$$\hat{\mu}_{i+1} = \frac{\sum X_s^> + \sum X_l^> + \sum X_h^> + n_l^< E[X_l^<] + n_h^< E[X_h^<]}{n_s + n_h} \quad (7.1)$$

$$\hat{\sigma}_{i+1}^2 = \frac{c_2}{n} \left[ \sum (X_s^> - \hat{\mu}_i)^2 + \sum (X_l^> - \hat{\mu}_i)^2 + \sum (X_h^> - \hat{\mu}_i)^2 + n_l^< E[(X_l^< - \hat{\mu}_i)^2] + n_h^< E[(X_h^< - \hat{\mu}_i)^2] \right] \quad (7.2)$$

$$\hat{\gamma}_{i+1} = \frac{c_3}{n \hat{\sigma}_{i+1}^3} \left[ \sum (X_s^> - \hat{\mu}_i)^3 + \sum (X_l^> - \hat{\mu}_i)^3 + \sum (X_h^> - \hat{\mu}_i)^3 + n_l^< E[(X_l^< - \hat{\mu}_i)^3] + n_h^< E[(X_h^< - \hat{\mu}_i)^3] \right] \quad (7.3)$$

where  $c_2$  and  $c_3$  are bias correction factors, defined as

$$c_2 = \frac{n_s + n_h^>}{n_s + n_h^> - 1} \quad (7.4)$$

$$c_3 = \frac{(n_s + n_h^>)^2}{(n_s + n_h^> - 1)(n_s + n_h^> - 2)} \quad (7.5)$$

and recalling  $n_s + n_h = n$ .

The expression  $E[X_h^<]$  is the expected value of an observation known to have a value less than the historical threshold  $X_h$ , and is a conditional expectation given that  $X < X_h$ , and is evaluated with

$$E[X|X \leq X_h; \hat{\tau}, \hat{\alpha}, \hat{\beta}] = \hat{\tau} + \hat{\beta} \frac{\Gamma\left(\frac{X_h - \hat{\tau}}{\hat{\beta}}, \hat{\alpha} + 1\right)}{\Gamma\left(\frac{X_h - \hat{\tau}}{\hat{\beta}}, \hat{\alpha}\right)} \quad (7.6)$$

where  $\Gamma(y, \alpha)$  is the incomplete gamma function

$$\Gamma(y, \alpha) = \int_0^y t^{\alpha-1} \exp(-t) dt. \quad (7.7)$$

DRAFT: August 31, 2017

---PROVISIONAL---  
THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.  
IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).  
IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

The expectation for higher order moments is

$$E[(X - \hat{\mu})^p | X \leq X_h; \hat{\tau}, \hat{\alpha}, \hat{\beta}] = \sum_{j=0}^p \binom{p}{j} \hat{\beta}^j (\hat{\tau} - \hat{\mu})^{p-j} \left[ \frac{\Gamma\left(\frac{X_h - \hat{\tau}}{\hat{\beta}}, \hat{\alpha} + j\right)}{\Gamma\left(\frac{X_h - \hat{\tau}}{\hat{\beta}}, \hat{\alpha}\right)} \right] \quad (7.8)$$

where  $p$  is the central moment index ( $p = 2, 3$ ). The conditional expectation for PILFs with  $X < X_l$  and threshold  $X_l$  are similar to equations 7.6 and 7.8.

The EMA moments shown in equations 7.1-7.3, and expected values shown in equations 7.6 and 7.8, utilize observations whose magnitudes are exactly known, where  $X_{lower} = X_{upper}$ . In the cases where flow magnitudes are described by intervals or binomial observations, these equations are modified to account for logarithms of flow intervals  $X_{Y,lower}$ ,  $X_{Y,upper}$  and are presented in equation 7.9. Information from broken, incomplete, and discontinued records, crest-stage gages, and multiple thresholds (for example, fig. 12) is easily represented by including additional expected value terms in the moments for each year  $Y$  or period where the flow interval or perception threshold varies.

The EMA employs the peak-flow intervals ( $Q_{lower}, Q_{upper}$ ) to estimate the moments of the LP-III distribution. Using base 10 logarithms of flows, where  $X_{lower} = \log_{10}(Q_{lower})$  and  $X_{upper} = \log_{10}(Q_{upper})$ , interval and binomial censored data are employed by replacing equation 7.8 with the following (Cohn and others, 1997):

$$E[(X - \hat{\mu})^p | X_{lower} \leq X \leq X_{upper}; \hat{\tau}, \hat{\alpha}, \hat{\beta}] = \sum_{j=0}^p \binom{p}{j} \hat{\beta}^j (\hat{\tau} - \hat{\mu})^{p-j} \left[ \frac{\Gamma\left(\frac{X_{upper} - \hat{\tau}}{\hat{\beta}}, \hat{\alpha} + j\right) - \Gamma\left(\frac{X_{lower} - \hat{\tau}}{\hat{\beta}}, \hat{\alpha} + j\right)}{\Gamma\left(\frac{X_{upper} - \hat{\tau}}{\hat{\beta}}, \hat{\alpha}\right) - \Gamma\left(\frac{X_{lower} - \hat{\tau}}{\hat{\beta}}, \hat{\alpha}\right)} \right]. \quad (7.9)$$

When information from a regional skew coefficient  $G$  is available, it is included directly in the EMA, ensuring that the adjusted mean and standard deviation fit the data. Equation 7.3 for the skew coefficient  $\hat{\gamma}_{i+1}$  is modified to include  $G$ , as

$$\hat{\gamma}_{i+1} = \frac{1}{(n + n_G) \hat{\sigma}_{i+1}^3} \left[ c_3 \left\{ \sum (X_s^> - \hat{\mu}_i)^3 + \sum (X_l^> - \hat{\mu}_i)^3 + \sum (X_h^> - \hat{\mu}_i)^3 + n_l^< E[(X_l^< - \hat{\mu}_i)^3] + n_h^< E[(X_h^< - \hat{\mu}_i)^3] \right\} + n_G G \hat{\sigma}_{i+1}^3 \right] \quad (7.10)$$

where  $n_G$  is the additional years of record assigned to the regional skew  $G$  (Griffis and others, 2004). A skew constraint is imposed on each EMA iteration so that  $\hat{\gamma}_{i+1} > -1.4$ , as it is unlikely that the population skew would be less than  $-1.4$ .

A general listing of computations for flood flow frequency using EMA, which are implemented in software (see the section Software and Examples), is as follows:

1. Check for low outliers with MGBT. If low outliers are detected, recode flows as censored data with an interval  $Q_{Y,lower} = 0$ ;  $Q_{Y,upper} = Q_{lower}$ . Adjust perception thresholds accordingly:  $T_{Y,lower} = Q_{lower}$ ;  $T_{Y,upper} = \infty$ .

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

2. Organize all flow intervals and perception thresholds for estimating parameters and confidence intervals. 1
3. Begin iterative fitting of the LP-III distribution using EMA with all data, including regional skew information. For each iteration, ensure that the weighted skew coefficient  $\tilde{G} \geq -1.41$  and the largest observation is within the fitted support of the distribution (for skews  $< 0$ ). 2  
3  
4
  - (a) Fit the LP-III with EMA, using at-site data, to estimate the at-site skew. 5
  - (b) Estimate the at-site skew coefficient MSE with EMA. 6
  - (c) Estimate a weighted skew coefficient. 7
  - (d) Fit the LP-III with EMA using a weighted skew coefficient. 8
  - (e) Test for convergence of EMA moments. If not converged, return to 3a. 9
4. Estimate quantile variances and compute confidence intervals based on the fitted LP-III model, including at-site and regional skew uncertainty. 10  
11

## The Generalized Expected Moments Algorithm

This section presents parameterizations of the P-III distribution and a generalized Expected Moments Algorithm (EMA). The notation and terms are utilized to explain uncertainty of EMA moments and confidence intervals. Bold terms, such as  $\mathbf{M}$  and  $\boldsymbol{\theta}$ , are used to indicate vectors or matrices. Carets (^) represent a sample estimate and tildes (~) indicate noncentral moments (on scalars) or estimators (on vectors). 13  
14  
15  
16

The P-III distribution is typically characterized by three parameters that correspond to location  $\{\tau\}$ , scale  $\{\beta\}$ , and shape  $\{\alpha\}$ , where the vector  $\boldsymbol{\theta} = \{\tau, \alpha, \beta\}$ . The P-III distribution is also characterized by noncentral moments  $\boldsymbol{\mu} = \{\tilde{\mu}_1, \tilde{\mu}_2, \tilde{\mu}_3\}$  (about zero) for algebraic tractability, and central moments  $\mathbf{M} = \{M, S, G\} = \{\mu, \sigma, \gamma\}$  for simplicity of explanation. 17  
18  
19  
20

Central moments are defined as 21

$$\begin{aligned} \mathbf{M} &= \begin{bmatrix} M \\ S \\ G \end{bmatrix} \equiv \begin{bmatrix} E[X] \\ \sqrt{E[(X-M)^2]} \\ E[(X-M)^3/S^3] \end{bmatrix} \equiv \begin{bmatrix} \tilde{\mu}_1 \\ \sqrt{\tilde{\mu}_2 - \tilde{\mu}_1^2} \\ \frac{\tilde{\mu}_3 - 3\tilde{\mu}_2\tilde{\mu}_1 + 2\tilde{\mu}_1^3}{\sqrt{\tilde{\mu}_2 - \tilde{\mu}_1^2}^3} \end{bmatrix} \\ &\equiv \begin{bmatrix} \tau + \alpha\beta \\ \sqrt{\alpha\beta^2} \\ \text{sign}(\beta)2/\sqrt{\alpha} \end{bmatrix}. \end{aligned} \quad (7.11)$$

Noncentral moments are 24

$$\begin{aligned} \tilde{\boldsymbol{\mu}} &\equiv \begin{bmatrix} \tilde{\mu}_1 \\ \tilde{\mu}_2 \\ \tilde{\mu}_3 \end{bmatrix} \equiv \begin{bmatrix} E_{\boldsymbol{\theta}}[X] \\ E_{\boldsymbol{\theta}}[X^2] \\ E_{\boldsymbol{\theta}}[X^3] \end{bmatrix} \equiv \begin{bmatrix} M \\ S^2 + M^2 \\ S^3G + 3S^2M + M^3 \end{bmatrix} \\ &= \begin{bmatrix} \alpha\beta + \tau \\ \alpha(1+\alpha)\beta^2 + 2\alpha\beta\tau + \tau^2 \\ \alpha(1+\alpha)(2+\alpha)\beta^3 + 3\alpha(1+\alpha)\beta^2\tau + 3\alpha\beta\tau^2 + \tau^3 \end{bmatrix}. \end{aligned} \quad (7.12)$$



1 The P-III distribution parameters are

$$2 \quad \theta = \begin{bmatrix} \alpha \\ \beta \\ \tau \end{bmatrix} = \begin{bmatrix} 4/G^2 \\ SG/2 \\ M - 2S/G \end{bmatrix} = \begin{bmatrix} \frac{4(\tilde{\mu}_2 - \tilde{\mu}_1^2)^3}{(\tilde{\mu}_3 - 3\tilde{\mu}_2\tilde{\mu}_1 + 2\tilde{\mu}_1^3)^2} \\ \frac{\tilde{\mu}_3 - 3\tilde{\mu}_2\tilde{\mu}_1 + 2\tilde{\mu}_1^3}{2(\tilde{\mu}_2 - \tilde{\mu}_1^2)} \\ \frac{\tilde{\mu}_3\tilde{\mu}_1 - 2\tilde{\mu}_2^2 + \tilde{\mu}_2\tilde{\mu}_1^2}{\tilde{\mu}_3 - 3\tilde{\mu}_2\tilde{\mu}_1 + 2\tilde{\mu}_1^3} \end{bmatrix}. \quad (7.13)$$

Here,  $E[\cdot]$  denotes the expectation operator. The formulas in equations 7.11, 7.12 and 7.13 facilitate converting one parametrization to another. When using sample estimates, the conversion from noncentral moments  $\hat{\mu}$  to central moments  $\mathbf{M}$  needs to include bias-correction factors with

$$\hat{\mathbf{M}} = \begin{bmatrix} 1 & \frac{N}{N-1} & \frac{\sqrt{N(N-1)}}{(N-2)} \end{bmatrix} * (\hat{\mu}). \quad (7.14)$$

3 A generalized EMA, employing central moments  $\mathbf{M}$ , is as follows, where  $N$  is the total record length. For  
4 convenience in the equations, the terms are abbreviated as  $X_{i,lower} = X_{i,l}$  and  $X_{i,upper} = X_{i,u}$ .

5 1. Initialize

- 6 (a) Set  $\hat{\mathbf{M}}_0 = \{0, 1, 0\}$   
7 (b) Define  $\varepsilon > 0$  as a satisfactory level of convergence.

8 A typical value for  $\varepsilon$  is  $10^{-10}$ .

9 2. Iterate: for  $j = 1, 2, \dots$

(a) Update expected moments

$$\hat{\mathbf{M}}_j = \begin{bmatrix} M_j \\ S_j^2 \\ G_j \end{bmatrix} = \begin{bmatrix} \frac{1}{N} \sum_{i=1}^N E_{\hat{\mathbf{M}}_{j-1}} [X_i | X_{i,l} \leq X_i < X_{i,u}] \\ \frac{1}{N-1} \sum_{i=1}^N E_{\hat{\mathbf{M}}_{j-1}} [(X_i - M_j)^2 | X_{i,l} \leq X_i < X_{i,u}] \\ \frac{N}{S^3(N-1)(N-2)} \sum_{i=1}^N E_{\hat{\mathbf{M}}_{j-1}} [(X_i - M_j)^3 | X_{i,l} \leq X_i < X_{i,u}] \end{bmatrix} \quad (7.15)$$

where

$$E_{\hat{\mathbf{M}}_{j-1}} [(X_i - M_j)^k | X_{i,l} \leq X_i < X_{i,u}] = \sum_{l=0}^k \binom{k}{l} E_{\hat{\mathbf{M}}_{j-1}} [X_i^l | X_{i,l} \leq X_i < X_{i,u}] (-M_j)^{k-l} \quad (7.16)$$

and, if the upper and lower bounds on  $X_i$  are equal (*i.e.*  $X_{i,l} = X_{i,u}$ , which means we know the exact value of  $X_i$ ), then

$$E_{\hat{\mathbf{M}}_{j-1}} [X_i^k | X_{i,l} \leq X_i < X_{i,u}] = X_{i,l}^k = X_{i,u}^k. \quad (7.17)$$

If  $X_{i,l} < X_{i,u}$ , then we have to evaluate the expectation

$$E_{\theta} [X^k | X_{i,l} \leq X < X_{i,u}] = \begin{cases} \sum_{j=0}^k \binom{k}{j} \beta^j \tau^{k-j} \left( \frac{\Gamma(\alpha+j, \frac{X_{i,u}-\tau}{\beta}, \frac{X_{i,l}-\tau}{\beta})}{\Gamma(\alpha, \frac{X_{i,u}-\tau}{\beta}, \frac{X_{i,l}-\tau}{\beta})} \right) & \beta < 0 \\ \sum_{j=0}^k \binom{k}{j} \beta^j \tau^{k-j} \left( \frac{\Gamma(\alpha+j, \frac{X_{i,l}-\tau}{\beta}, \frac{X_{i,u}-\tau}{\beta})}{\Gamma(\alpha, \frac{X_{i,l}-\tau}{\beta}, \frac{X_{i,u}-\tau}{\beta})} \right) & \beta > 0 \end{cases} \quad (7.18)$$

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

where  $\theta$  is the P-III parameters corresponding to  $\tilde{\mathbf{M}}_{j-1}$ , and

$$\Gamma(\alpha, X_{i,l}, X_{i,u}) = \int_{\max(0, X_{i,l})}^{\max(0, X_{i,u})} t^{\alpha-1} \exp(-t) dt. \quad (7.19)$$

(b) If available, weight with regional skew. This can be done conceptually via:

$$\tilde{G}_j = \frac{MSE_G \hat{\gamma}_j + MSE_{\hat{\gamma}} G}{MSE_G + MSE_{\hat{\gamma}}} \quad (7.20)$$

and the number of years are used as weights, as in equation 7.10.

(c) Test for convergence. If  $||\hat{\mathbf{M}}_j - \hat{\mathbf{M}}_{j-1}|| < \epsilon$ , return  $\mathbf{M} = \hat{\mathbf{M}}_j$  as the EMA estimate. Otherwise, increment  $j$  and return to 2a.

## Uncertainty of EMA Moments

Uncertainty of moments, specifically the at-site skew coefficient ( $\hat{\gamma}$ ), are estimated with EMA. Details and equations are presented in appendix A1 of Cohn and others (2001) and in Cohn (2015).

For cases where there is historical information, PILFs, a gage-base discharge, or some type of censored or interval data, EMA utilizes an approach to estimate  $MSE_{\hat{\gamma}}$  that is based on all the data. This includes censored data, intervals, historical information, and PILFs, including the P-III distribution parameters, as they are used in estimating the moments with EMA. For convenience in the equations, the terms are abbreviated as  $X_{i,lower} = X_{i,l}$  and  $X_{i,upper} = X_{i,u}$ . Conceptually, this is done as follows:

$$MSE_{\hat{\gamma}} \approx \text{VAR}(\hat{\gamma}) \approx \text{VAR}(\hat{m}_3) \approx \frac{1}{n} f(X_{i,l}, X_{i,u}, T_{i,l}, T_{i,u}, \hat{\theta}) \quad (7.21)$$

where  $MSE_{\hat{\gamma}}$  is proportional to the variance (VAR) of the skew ( $\hat{\gamma}$ ), is proportional to the variance of the third noncentral moment  $\text{VAR}(\hat{m}_3)$ , is a function ( $f$ ) of the observations (including censored data), and P-III parameters  $\hat{\theta}$ . In this case,  $n$  is the total record length (for example,  $n = n_s + n_h = N$ ), including any historical period, PILFs, and censored and interval data.

As presented in Appendix A1 of Cohn and others (2001), EMA estimates the variance of each of the noncentral moments  $\{\hat{\mu} = [\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3]\}$ , where  $\{\hat{\mu} = \hat{\mathbf{M}} = [m_1, m_2, m_3]\}$  as outlined in Cohn and others (2001). The noncentral moment  $\hat{m}_3$  (moment computed around zero) is used to estimate  $MSE_{\hat{\gamma}}$ . Key equations from Cohn and others (2001, appendix A1) are presented below.

The EMA estimates noncentral moments  $\hat{\mathbf{M}} = [\hat{m}_1, \hat{m}_2, \hat{m}_3]$  that directly take into account censored data, via

$$\hat{\mathbf{M}} = (1/n) \sum_{i=1}^n \mathbf{x}(\psi(X_i), \hat{\mathbf{M}}) \mathcal{J}[\psi(X_i)] \quad (7.22)$$

where

$$\mathcal{J}[X] \equiv \begin{bmatrix} \mathcal{J}(X < a) \\ \mathcal{J}(a \leq X \leq b) \\ \mathcal{J}(X > b) \end{bmatrix} \quad (7.23)$$

$$\mathcal{J}(\text{condition}) \equiv \begin{cases} 1 & \text{condition} = \text{true} \\ 0 & \text{otherwise} \end{cases} \quad (7.24)$$

DRAFT: August 31, 2017

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

and

$$\chi(\psi(X), \mathbf{M}) = \begin{bmatrix} E_{\theta[\mathbf{M}]}[X|X < a] & X_i & E_{\theta[\mathbf{M}]}[X|X > b] \\ E_{\theta[\mathbf{M}]}[X^2|X < a] & X_i^2 & E_{\theta[\mathbf{M}]}[X^2|X > b] \\ E_{\theta[\mathbf{M}]}[X^3|X < a] & X_i^3 & E_{\theta[\mathbf{M}]}[X^3|X > b] \end{bmatrix}. \quad (7.25)$$

1 The function  $\mathcal{J}[X]$  defines the censored data category for the flow logarithms  $X$ . The following three categories  
 2 are used: “less,” where  $X$  is less than the “perception threshold”  $a$ ; “between,” where  $X$  is within the closed  
 3 interval  $[a, b]$ ; or “greater” if  $X$  is known to exceed some “perception threshold”  $b$  (table 7.1). These threshold  
 4 categories  $[a, b]$  correspond to those described in the section Data Representation using Flow Intervals and  
 5 Perception Thresholds, where “between” is the “interval” category and “greater” is the “binomial” category.  
 6 The “less than” category covers unobserved historical floods, flows less than a gage base, or low outliers.  
 7 The magnitude of  $X$  is known if  $X$  is within  $[a, b]$ . Only a threshold on  $X$  can be identified if  $X < a$  or  
 8  $X > b$ . The number of observations in each of these categories is a random variable, denoted  $n_l, n_b$ , and  
 9  $n_g$ , respectively. Because each  $X$  must fall into one of the three categories, the total sample size  $n$  is constant,  
 10 where  $n = n_l + n_b + n_g$ .

The  $MSE_{\hat{\gamma}}$  can be estimated by taking the variance of equation 7.22, as in equation 7.26. The formula for the asymptotic variance of the EMA moments estimator, denoted  $\tilde{\Sigma}_{\hat{\mu}}$ , is derived in Cohn and others (2001, appendix A1). It is obtained by linearizing the expectations in equation 7.22 and solving for  $\mathbf{M}$  in terms of the sample  $X_i$  values. The estimator  $\tilde{\Sigma}_{\hat{\mu}}$  is then expressed as a function of the population parameters, the record lengths, and the censoring thresholds. It can be used as an **estimator** of the variance-covariance matrix given estimated parameters ( $\hat{\Sigma}_{\mathbf{M}}$ ):

$$VAR \hat{\mathbf{M}} = VAR \begin{bmatrix} \hat{m}_1 \\ \hat{m}_2 \\ \hat{m}_3 \end{bmatrix} \approx \tilde{\Sigma}_{\hat{\mu}} \approx \hat{\Sigma}_{\mathbf{M}}. \quad (7.26)$$

The variance of  $\hat{\mathbf{M}}$  is (Cohn and others, 2001, eq. 55):

$$\tilde{\Sigma}_{\hat{\mu}} = \frac{1}{n^2} \mathbf{A} (Var[\mathbf{B}] + Var[\mathbf{C}]) \mathbf{A}' \quad (7.27)$$

**Table 7.1.** Expected Moments Algorithm censored-data threshold categories.

Value of $X_i$	Category	$(T_{i,l}, T_{i,u})$
$x < a$	$l$	$(-\infty, a)$
$a \leq X \leq b$	$b$	$(X, X)$
$x > b$	$g$	$(b, \infty)$

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

where

$$\begin{aligned}\mathbf{B} &= \boldsymbol{\mu}_{\mathbf{X}} \mathbf{n} \\ \mathbf{C} &= \sum_{i=1}^{\mu_{n_b}} (\mathbf{X}_i - \boldsymbol{\mu}_{\mathbf{X}_b}) \\ \mathbf{D} &= \frac{\mu_{n_l} \mathbf{J}_l + \mu_{N_g} \mathbf{J}_g}{n} \\ \mathbf{A} &= (\mathbf{I} - \mathbf{D})^{-1}\end{aligned}\quad (7.28)$$

and  $\boldsymbol{\mu}_{\mathbf{X}}$  is the vector of expected values for noncentral moments given parameters and value of  $X$  (Cohn and others, 2001, eq. 50–51). The variance of  $\mathbf{B}$  is given by

$$\text{Var}[\mathbf{B}] = \boldsymbol{\mu}_{\mathbf{X}} \text{Var}[\mathbf{n}] \boldsymbol{\mu}_{\mathbf{X}}'. \quad (7.29)$$

The large-sample variance of  $\mathbf{C}$  is the expected value of the number of terms multiplied by the variance of each term:

$$\text{Var}[\mathbf{C}] = \mu_{n_b} \begin{bmatrix} V_{1,1} & V_{1,2} & V_{1,3} \\ V_{2,1} & V_{2,2} & V_{2,3} \\ V_{3,1} & V_{3,2} & V_{3,3} \end{bmatrix}. \quad (7.30)$$

The MSE of the EMA at-site skewness coefficient is estimated using a first-order approximation (Cohn and others, 2001, eq. 55), reproduced above as equation 7.27, with  $\hat{m}_3$  as the noncentral moment of interest.

## Confidence Intervals with EMA

A simple formula for a confidence interval on a flood quantile  $\hat{X}_q$  is (Stedinger and others, 1993; Cohn and others, 2001):

$$\hat{X}_q \pm z_{1-\alpha/2} \sqrt{\text{var}(\hat{X}_q)} \quad (7.31)$$

where  $q$  is the quantile of interest (e.g.  $q = 0.99$ ),  $z_{1-\alpha/2}$  is the  $(1 - \alpha)/2$  quantile of the standard Normal distribution,  $\alpha$  is the confidence level and

$$\sqrt{\text{var}(\hat{X}_q)} = \hat{\sigma}_{\hat{X}_q} \quad (7.32)$$

is the estimated standard error of the flood quantile. Typically, the confidence level  $\alpha = 0.05$ , resulting in a 90-percent confidence interval (5-percent and 95-percent confidence limits).

Confidence intervals for flood quantiles ( $\hat{X}_p$ ) are estimated with EMA. Cohn and others (2001) derive EMA confidence intervals in detail and provide key equations. Cohn (2015) improve the EMA confidence intervals for skews  $|\hat{\gamma}| > 0.5$ .

Confidence intervals are estimated using

$$\left( \hat{X}_p + \frac{\hat{\sigma}_{\hat{X}_p} T_{v,(1-\varepsilon)/2}}{1 - \kappa T_{v,(1-\varepsilon)/2}}, \hat{X}_p + \frac{\hat{\sigma}_{\hat{X}_p} T_{v,(1+\varepsilon)/2}}{1 - \kappa T_{v,(1+\varepsilon)/2}} \right) \quad (7.33)$$

where  $T_v$  is a Student's  $T$  variate (Abramowitz and Stegun, 1964),  $\varepsilon$  is the confidence level,  $\hat{\sigma}_{\hat{X}_p}$  is the standard

## 104 Guidelines for Determining Flood Flow Frequency – Bulletin 17C

deviation of the quantile  $\hat{X}_p$ , and

$$\kappa \equiv \frac{\hat{Cov}[\hat{X}_p, \hat{\sigma}_{\hat{X}_p}]}{\hat{\sigma}_{\hat{X}_p}^2} \quad (7.34)$$

- 1 is a function of the sample size and the censoring threshold (and, to some extent, of  $\alpha$ ). Estimators for  
 2  $\hat{Cov}[\hat{X}_p, \hat{\sigma}_{\hat{X}_p}]$  and  $\hat{\sigma}_{\hat{X}_p}^2$  are available from Cohn and others (2001, eq. 70).

The asymptotic variance of  $\hat{X}_p$  can be obtained from a first-order expansion of  $\hat{X}_p$  as a function of  $\mathbf{M}$ :

$$\hat{X}_p \approx X_p + \mathbf{J}_{\hat{X}_p}(\mathbf{M} - \boldsymbol{\mu}_{\mathbf{M}}) \quad (7.35)$$

where

$$\mathbf{J}_{\hat{X}_p} = \left[ \frac{\partial \hat{X}_p}{\partial \hat{m}_1} \quad \frac{\partial \hat{X}_p}{\partial \hat{m}_2} \quad \frac{\partial \hat{X}_p}{\partial \hat{m}_3} \right]. \quad (7.36)$$

- 3 The Jacobian can be evaluated by first computing derivatives with respect to  $\{\alpha, \beta, \tau\}$  and then applying the  
 4 chain rule.

The variance of  $\hat{X}_p$  can be approximated by

$$\tilde{\sigma}_{\hat{X}_p}^2 \approx \mathbf{J}_{\hat{X}_p} \cdot \tilde{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\mu}}} \cdot \mathbf{J}_{\hat{X}_p}' \quad (7.37)$$

- 5 where the linearized standard deviation,  $\tilde{\sigma}_{\hat{X}_p}$ , is defined as  $\sqrt{\tilde{\sigma}_{\hat{X}_p}^2}$ .

- 6 Cohn (2015) provides improved estimates of  $\text{Var}[\hat{X}_p]$  using inverse quadrature.

DRAFT: August 31, 2017

---PROVISIONAL---  
 THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
 PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
 PEER REVIEW PLAN.  
 IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
 GEOLOGICAL SURVEY (USGS).  
 IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
 TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

## Appendix 8. Record Extension with Nearby Sites

---

1

2

DRAFT: August 31, 2017

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.



DRAFT: August 31, 2017

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

## Record Extension with Nearby Sites

This appendix describes the background, guidance, computational details, and an example for record extension. Record extension uses information from and computational details for transfer of information from a long-record nearby site to a site of concern which has a relatively shorter record. The transfer is based upon the observed correlation between the base-10 logarithms of the annual maximum series available at the two sites. This is an attractive method to transfer information from a long-record site to the site of interest while honoring the PILF criterion and the relative information provided by historical information and regional skewness estimators. An example illustrates the method.

Formally, such a method based upon the cross correlation between flood peaks at a short- and long-record site is called Record Augmentation (see Glossary for definitions). However, because the Bulletin 17C algorithm uses a PILF threshold to restrict the frequency analysis to floods greater than the PILF threshold, a simple record augmentation procedure that ignores that restriction would not be appropriate. The supplemental information provided by the record augmentation analysis is represented as an extension of the original flood record. The additional length of record is selected to approximate the “effective record length” of the additional information for the variance developed for the flood frequency analysis at the short-record site. The PILF threshold can then be applied to this extended record to allow for a consistent analysis of flood peaks that are greater than that threshold.

Record augmentation is attractive when two conditions are met. First, the record available at the site of interest is relatively short. And, there is a nearby longer flood record site to extend the time series at a short record site, when whose flood series is highly correlated with the flood series at the short-record site. Then the longer record can be used to effectively extend the record at the short-record site employing the cross-correlation between floods at the two sites is high ( $\hat{\rho} > 0.80$ ). It is recommended that record extension be considered when the short. The recommendation is that record augmentation be considered whenever the cross-correlation  $\rho > 0.80$  and the short-record site is less than 20 years, with a minimum length of 10 years.

. It may be useful even when the short-record site exceeds 20 years. In any case, a minimum of 7 or 8 years concurrent years of record are needed at the two sites considered. Matalas and Jacobs (1964) and Vogel and Stedinger (1985) discuss the record augmentation procedures, and provide more complex criteria for when the cross-correlation between the two flood series and the concurrent record length are insufficient for the procedures to result in a gain in the mean and variance at the short-record site. The larger the cross correlation the better.

## Record Augmentation of the Mean and Variance

Matalas and Jacobs (1964) developed an approach for obtaining unbiased estimates of the mean and variance of the lengthened time series (observed plus extended record) using based upon ordinary least squares (OLS) regression without adding a random noise term. This approach is the basis of the “Two Station Comparison” method that is described in Bulletin 17B (IACWD, 1982, Appendix 7). In (IACWD, 1982, appendix 7), that method Using the methods of moments recommended in Bulletin 17B, improved estimates of the mean and variance (standard deviation) are obtained but no additional years of record are estimated. In order to be useful in flood frequency with, annual peak-flow estimates are needed, rather than improved estimates of the flow statistics (mean and variance).

The Maintenance of Variance Extension (MOVE) techniques, as described by Hirsch (1982), are useful approaches for extending flood data in time, based on records from a nearby site. As implied by the name, these techniques maintain the variance of the estimated data unlike OLS regression. If an ensemble of points is estimated from OLS regression, the estimated values will have lesser variability than the true or population

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

values unless a random noise term is added. However, adding a random noise term to the regression estimates does not achieve a unique extended record. Therefore, the loss of variance from regression analysis is a problem if the estimated values are used in subsequent statistical analyses such as frequency analyses.

Hirsch (1982) described two MOVE techniques, MOVE.1 and MOVE.2. The MOVE procedures are based on only one independent variable and the assumption is that there is a linear relation between the dependent and independent variables. If the annual peak flows are not linearly related, then it is common to transform the data using a logarithmic transformation because the logarithms of the flows tend to be linearly related. The MOVE techniques are another way of fitting a linear relation to data similar to OLS regression.

The MOVE.1 technique described by Hirsch (1982) is the simplest and uses the  $n_1$  years of concurrent record at the two sites. Only the means and standard deviations for  $y_i$  and  $x_i$  for the concurrent record are used to define the MOVE.1 relation. The MOVE.2 technique described by Hirsch (1982) utilizes the can be computed based upon the Matalas-Jacobs estimators (Matalas and Jacobs, 1964) for the mean and variance of  $y_i$  at the short-term station plus the additional years of record  $x_{n_1+1}$  to  $x_{n_1+n_2}$  at the long-term station that were not observed at the short-term station. The MOVE.1 and MOVE.2 techniques ensure that procedure and then used in the moments of the historical sequence ( $y_1$  to  $y_{n_1}$ ) are preserved.

Vogel and Stedinger (1985) suggested several variations on the MOVE techniques presented by Hirsch (1982). Their MOVE.3 approach provides an estimate of the mean and variance for the short computations. Moran (1974) considered use of more than one explanatory variable – a potentially useful extension of the concept. Because Bulletin 17C introduces the PILF criteria with EMA, integration of regional skewness information, and potentially historical information and other data, adopting information provided by record augmentation is not as easy. That will be accomplished with an extended systematic record that represents the additional information provided by correlation with the long-record site that is correct for the complete record, if the correlation coefficient exceeds a critical value. The MOVE.3 approach is based on a linear relation that ensures the mean and variance of the lengthened sequence (observed plus extended record) ( $y_1$  to  $y_{n_1+n_2}$ ) will be preserved and equal to the Matalas and Jacobs (1964) estimators (Vogel and Stedinger, 1985). Each successive MOVE technique provides slightly more accurate estimates. The recommended approach for record extension is MOVE.3. It is called ‘MOVE’ in the remainder of this appendix and The extended systematic record is described below, introduced into the EMA computations, and then is subject to the PILF criteria.

Consider a peak-flow time series at a Let the base 10 logarithms of floods  $y = \log_{10}(Q_y)$  at the short-record site  $y_1, \dots, y_{n_1}$  with at least 10 years of data. Information from a peak-flow time series at a hydrologically of interest be denoted  $y_i$ , and logarithms of floods  $x = \log_{10}(Q_x)$  at the long-relevant longer record site  $x_1, \dots, x_{n_1}, x_{n_1+1}, \dots, x_{n_1+n_2}$ , with similar climatic and watershed characteristics, is used to extend the record at site  $y$ , where  $y_i$  and record site be denoted  $x_i$  are the base 10 logarithms of flows  $Q$  at sites  $y$  and  $x$ . Furthermore let  $n_1$  is be the length of the short record, and  $n_1 + n_2$  is the length of the long record. Note so that the series is  $y_i$ , for  $i = 1, \dots, n_1$ . Let  $n_2$  be the additional years of record at the  $x$ -site, so that series can be written  $x_i$ , for  $i = 1, \dots, n_1 + n_2$ . For convenience, the presentation assumes that the  $n_1$  concurrent observations between the two sites  $y$  and  $x$  do not need to correspond to the first  $n_1$  observations, nor do they. This need not be the case nor do the concurrent values need to be consecutive. Here the  $x$  series should be from a hydrologically-relevant site with similar climatic and watershed characteristics.

A linear regression model is used to extend the record at the short site  $y_i$  (Vogel and Stedinger, 1985):

$$\hat{y}_i = a' + b(x_i - \bar{x}_2)$$

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

where  $\bar{x}_2$  is the mean of the longer-term station for the non-overlapping period, estimated from

$$\bar{x}_2 = \frac{1}{n_2} \sum_{i=n_1+1}^{n_1+n_2} x_i$$

Definitions employed in the presentation include these sample means and variances ~~the intercept  $a'$  is~~

$$\underline{n_2}$$

~~the slope is estimated from~~

$$b^2 = \frac{(n_1 + n_2 - 1)\hat{\sigma}_y^2 - (n_1 - 1)s_{y1}^2 - n_1(\bar{y}_1 - \hat{\mu}_y)^2 - n_2(a' - \hat{\mu}_y)^2}{(n_2 - 1)s_{x2}^2}$$

and  $\hat{\mu}_y$  and  $\hat{\sigma}_y^2$  are the unbiased estimators (Matalas and Jacobs, 1964) for the mean and variance of the complete extended record. The sample statistics  $(\bar{y}_1, \bar{x}_1, s_{y1}^2, s_{x1}^2)$  of the concurrent records in equations 8.23 and 8.24 are estimated as follows:-

$$\bar{y}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} y_i \quad (8.1)$$

$$\bar{x}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} x_i \quad (8.2)$$

$$\bar{x}_2 = \frac{1}{n_2} \sum_{i=n_1+1}^{n_1+n_2} x_i \quad (8.3)$$

$$s_{y1}^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (y_i - \bar{y}_1)^2 \quad (8.4)$$

$$s_{x1}^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_i - \bar{x}_1)^2 \quad (8.5)$$

and

$$s_{x2}^2 = \frac{1}{n_2 - 1} \sum_{i=n_1+1}^{n_1+n_2} (x_i - \bar{x}_2)^2. \quad (8.6)$$

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

## 110 Guidelines for Determining Flood Flow Frequency – Bulletin 17C

~~The~~ With these definitions, the Matalas-Jacobs estimators are (Matalas and Jacobs, 1964; Vogel and Ste-dinger, 1985):

$$\hat{\mu}_y = \bar{y}_1 + \frac{n_2}{n_1 + n_2} \hat{\beta} (\bar{x}_2 - \bar{x}_1) \quad (8.7)$$

and

$$\hat{\sigma}_y^2 = \frac{1}{n_1 + n_2 - 1} \left[ (n_1 - 1) s_{y_1}^2 + (n_2 - 1) \hat{\beta}^2 s_{x_2}^2 + (n_2 - 1) \alpha^2 (1 - \hat{\rho}^2) s_{y_1}^2 + \frac{n_1 n_2}{(n_1 + n_2)} \hat{\beta}^2 (\bar{x}_2 - \bar{x}_1)^2 \right] \quad (8.8)$$

where

$$\hat{\rho} = \hat{\beta} \frac{s_{x_1}}{s_{y_1}} \quad (8.9)$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n_1} (x_i - \bar{x}_1)(y_i - \bar{y}_1)}{\sum_{i=1}^{n_1} (x_i - \bar{x}_1)^2} \quad (8.10)$$

and

$$\alpha^2 = \frac{n_2(n_1 - 4)(n_1 - 1)}{(n_2 - 1)(n_1 - 3)(n_1 - 2)}. \quad (8.11)$$

~~Record extension is an appropriate technique when there is substantial improvement to both~~ It is important to understand the precision of the Matalas-Jacobs estimators of the mean and variance of flood peaks at the short-record site. Using the definitions introduced above, the short-record site variance of the estimated mean at the y-site is

$$\text{var}(\hat{\mu}_y) = \frac{\sigma_y^2}{n_1} \left[ 1 - \frac{n_2}{n_1 + n_2} \left( \rho^2 - \frac{1 - \rho^2}{n_1 - 3} \right) \right] \quad (8.12)$$

whereas the variance of the variance of floods at the y-site is

$$\text{var}(\hat{\sigma}_y^2) = \frac{2\sigma_y^4}{n_1 - 1} + \frac{n_2 \sigma_y^4}{(n_1 + n_2 - 1)^2 (n_1 - 3)} (A \rho^4 + B \rho^2 + C) \quad (8.13)$$

where

$$A = \frac{(n_2 + 2)(n_1 - 6)(n_1 - 8)}{(n_1 - 5)} + (n_1 - 4) \left( \frac{n_1 n_2 (n_1 - 4)}{(n_1 - 3)(n_1 - 2)} - \frac{2n_2 (n_1 - 4)}{(n_1 - 3)} - 4 \right) \quad (8.14)$$

$$B = \frac{6(n_2 + 2)(n_1 - 6)}{(n_1 - 5)} + 2(n_1^2 - n_1 - 14) + (n_1 - 4) \left( \frac{2n_2 (n_1 - 5)}{(n_1 - 3)} - 2(n_1 + 3) - \frac{2n_1 n_2 (n_1 - 4)}{(n_1 - 3)(n_1 - 2)} \right) \quad (8.15)$$

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

$$C = 2(n_1 + 1) + \frac{3(n_2 + 2)}{(n_1 - 5)} - \frac{(n_1 + 1)(2n_1 + n_2 - 2)(n_1 - 3)}{(n_1 - 1)} + (n_1 - 4) \left( \frac{2n_2}{(n_1 - 3)} + 2(n_1 + 1) + \frac{n_1 n_2 (n_1 - 4)}{(n_1 - 3)(n_1 - 2)} \right) \quad (8.16)$$

and these correspond to equations in Matalas and Jacobs (1964), IACWD (1982, appendix 7), and Vogel and Stedinger (1985)

2

If the improvement in the precision of the estimate of the  $y$ -mean is described by an increase in the effective record length  $n_e$ , then  $n_e$  would be defined by the equation

$$\text{var}(\hat{\mu}_y) = \frac{\sigma_y^2}{(n_e + n_1)} \quad (8.17)$$

where  $\sigma_y^2$  is the true variance of the  $y$ -series. Because  $\text{var}(\hat{\mu}_y)$  in equation 8.12 is proportional to  $\sigma_y^2$ , based on the longer concurrent record. This occurs when the variance of the combined record  $\hat{\sigma}_y^2$  is less than the variance of the short record  $s_{y1}^2$ . This is typically the case when  $\hat{\rho} > 0.80$  (Vogel and Stedinger, 1985).

$$\rho^2 > \frac{-B \pm (B^2 - 4AC)^{1/2}}{2A}$$

where  $y$  series cancels out in the computation of  $n_e$ . Thus the total effective record length in terms of the estimated mean is

$$n_e + n_1 = \frac{n_1}{\left[ 1 - \frac{n_2}{n_1 + n_2} \left( \rho^2 - \frac{1 - \rho^2}{n_1 - 3} \right) \right]} \quad (8.18)$$

and corresponds to equation 7-7 in IACWD (1982, appendix 7).

$$A = \frac{(n_2 + 2)(n_1 - 6)(n_1 - 8)}{(n_1 - 5)} + (n_1 - 4) \left( \frac{n_1 n_2 (n_1 - 4)}{(n_1 - 3)(n_1 - 2)} - \frac{2n_2(n_1 - 4)}{(n_1 - 3)} - 4 \right)$$

In terms of the computation of flood quantiles, the estimator of the variance is much more important than the estimator of the mean. As a result, in the use of the Matalas-Jacobs estimators for flood frequency analysis, the effective increase in the record length in terms of the precision of the variance is the critical issue. Using equations 8.13–8.16, the needed relationship for the total effective record length for the estimated variance is

$$\frac{2}{n_1 - 1} + \frac{n_2}{(n_1 + n_2 - 1)^2 (n_1 - 3)} (\underline{n_1 A} \rho^4 + \underline{3B} \rho^2 + \underline{C}) - \frac{2n_1 n_2 (n_1 - 4)}{(n_1 - 3)(n_1 - 2)} + 1.$$

3

---PROVISIONAL---

(8.19)

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017



$$C = 2(n_1 + 1) + \frac{3(n_2 + 2)}{(n_1 - 5)} - \frac{(n_1 + 1)(2n_1 + n_2 - 2)(n_1 - 3)}{(n_1 - 1)} + (n_1 - 4) \left( \frac{2n_2}{(n_1 - 3)} + 2(n_1 + 1) + \frac{n_1 n_2 (n_1 - 4)}{(n_1 - 3)(n_1 - 2)} \right)$$

and thereby assures

Use of the Matalas-Jacobs estimators is appropriate when they result in a substantial improvement in both the mean and variance estimators of flood peaks at the short-record  $y$  site, corresponding to an  $n_e$  of 4 or 5 or more. This is typically the case when  $\hat{\rho} > 0.70$ ,  $\hat{\rho} > 0.80$  (Vogel and Stedinger, 1985). The above equations were developed under the assumption that the logarithms of concurrent flow observations at the short-record site  $y_i$  and long-record site  $x_i$  have a joint normal probability distribution with a skewness of zero. When this assumption is seriously violated, the above equations are not exact and this technique should be used with caution.

For a short-record site where record extension is deemed valuable, annual peak flow estimates for an “extended” record are needed for use with and the recommended procedures described in Section . This “extended” record is comprised of the short

## Record Extension with MOVE

Having computed the improved Matalas-record observations  $n_1$  plus the “extended” observations denoted  $n_e$ , to make a longer record  $n_1 + n_e$ . To accomplish this, the Matalas-Jacobs-Jacobs estimators of the mean and variance (equations 8.7 and 8.8) that are preserved in the longer record  $n_1 + n_e$ , are used with MOVE to estimate the for the  $y$ -site, and effective records  $n_1 + n_e$  corresponding to the improvement in the variance estimator (eqn. 8.19), the next step is to generate  $n_e$  flows. The extended record length  $n_1 + n_e$  is estimated using the variance of the improved variance (8.8), and the  $n_1 + n_e$  flows are then used in . In this way, appropriate confidence intervals can be estimated.

The extended record length  $n_1 + n_e$  can be estimated by using the equivalent record length of the improved variance of the Matalas-Jacobs estimators. can be estimated by using the equivalent record length of the variance of the Matalas-Jacobs estimators (eqn placeholder for mean ):-

$$n_1 + n_e = \frac{n_1}{1 - \frac{n_2}{n_1 + n_2} (\hat{\rho}^2 - \frac{1 - \hat{\rho}^2}{n_1 - 3})}$$

where ( This equation is a placeholder equn for neff based on the mean from B17B equn 7-7. We need to use the variance ratio from page 717 of Vogel/Stedinger 1985 equns 10 and 11 – do I have this correct?)- additional observations to be added to the  $y$ -record to introduce this additional information into the frequency analysis. This is done employing the Maintenance of Variance Extension (MOVE) idea described by Hirsch (1982). As implied by the name, these techniques were developed to maintain the variance of a generated  $y$ -series when it was extended using a supplemental longer series, denoted here as  $x$ . As suggested by Vogel and Stedinger (1985), one can be careful in how model parameters are selected so that the original and extended values have exactly the mean and variance that a hydrologist specifies. In our case, the original  $n_1$  observations represent themselves, and the additional  $n_e$  values represent the information provided by the additional observations in the  $x$ -series, given the less than perfect cross-correlation between the two series. It should be clear that the procedure adds just  $n_e$  observations to the original  $y$  series and those  $n_e$  observations have the values needed to transfer information about the mean and variance at the  $y$ -site.

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

Estimation of the equivalent record length can be accomplished by using the mean or the variance. The  $n_1 + n_e$  estimate is derived using the mean. A linear regression model is used to extend the record at the short site  $y_i$  by  $n_e$  (Vogel and Stedinger, 1985):

$$\hat{y}_i = a + b(x_i - \bar{x}_e) \quad i = n_1 + 1, \dots, n_1 + n_e \quad (8.20)$$

where  $\bar{x}_e$  is the mean of the extended record  $\hat{\mu}_y$  as follows:

$$\text{var}(\hat{\mu}_y) = \frac{\hat{\sigma}_y^2}{(n_e + n_1)}$$

$$n_e + n_1 = \frac{\hat{\sigma}_y^2}{\text{var}(\hat{\mu}_y)}$$

$$\text{var}(\hat{\mu}_y) = \frac{s_{y1}^2}{n_1} \left[ 1 - \frac{n_2}{n_1 + n_2} (\hat{\rho}^2 - \frac{1 - \hat{\rho}^2}{n_1 - 3}) \right]$$

$$n_e + n_1 = \frac{\hat{\sigma}_y^2}{\frac{s_{y1}^2}{n_1} \left[ 1 - \frac{n_2}{n_1 + n_2} (\hat{\rho}^2 - \frac{1 - \hat{\rho}^2}{n_1 - 3}) \right]}$$

Assume  $\hat{\sigma}_y^2 \approx s_{y1}^2$ . Then the equivalent years of record for the mean is

$$n_e + n_1 = \frac{n_1}{\left[ 1 - \frac{n_2}{n_1 + n_2} (\hat{\rho}^2 - \frac{1 - \hat{\rho}^2}{n_1 - 3}) \right]}$$

and corresponds to equation (7-7) in IACWD (1982, Appendix 7).

1

Using a similar procedure as the mean, the equivalent years using the variance of the extended record  $\hat{\sigma}_y$  is estimated as follows.  $n_e$  x-series values from the non-overlapping period used to extend the y series. That is, given

$$\text{var}(\bar{x}_e) = \frac{2\hat{\sigma}_y^4}{(n_e + n_1 - 1)} \frac{1}{n_e} \sum_{i=n_1+1}^{n_1+n_e} x_i \quad (8.21)$$

$$\text{var}(\hat{\sigma}_y^2)$$

From Matalas and Jacobs (1964) and Vogel and Stedinger (1985): and

$$\text{var}(s_{x_e}^2) = \frac{2s_{y1}^4}{n_1 - 1} + \frac{n_2 s_{y1}^4}{(n_1 + n_2 - 1)^2 (n_1 - 3)} \frac{1}{n_e - 1} \sum_{i=n_1+1}^{n_1+n_e} (A^4 + B^2 + C x_i - \bar{x}_e)^2. \quad (8.22)$$

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

where-

$$A = \frac{(n_2 + 2)(n_1 - 6)(n_1 - 8)}{(n_1 - 5)} + (n_1 - 4) \left( \frac{n_1 n_2 (n_1 - 4)}{(n_1 - 3)(n_1 - 2)} - \frac{2n_2 (n_1 - 4)}{(n_1 - 3)} - 4 \right)$$

$$B = \frac{6(n_2 + 2)(n_1 - 6)}{(n_1 - 5)} + 2(n_1^2 - n_1 - 14) + (n_1 - 4) \left( \frac{2n_2 (n_1 - 5)}{(n_1 - 3)} - 2(n_1 + 3) - \frac{2n_1 n_2 (n_1 - 4)}{(n_1 - 3)(n_1 - 2)} \right)$$

The intercept  $a$  to use to generate  $n_e$  new values is computed as

$$n_1 + 1) + \frac{n_1 n_2 (n_1 - 4)}{(n_1 - 3)(n_1 - 2)} e$$

(8.23)

Inserting equation 8.13 into equation ?? and assuming  $\hat{\sigma}_y^4 \approx s_{y1}^4$ , then one obtains the equivalent record length for the improved variance as with the slope ( $b > 0$ ) estimated from

$$b^2 = \frac{(n_1 + n_e - 1)\hat{\sigma}_y^2 - (n_1 - 1)s_{y1}^2 - n_1(\bar{y}_1 - \hat{\mu}_y)^2 - n_e(a - \hat{\mu}_y)^2}{(n_e - 1)s_{xe}^2}$$

$$\frac{2}{n_1 - 1} + \frac{n_2}{(n_1 + n_2 - 1)^2(n_1 - 3)} (A\hat{\rho}^4 + B\hat{\rho}^2 + C) + 1.$$

(8.24)

Should  $b^2$  be less than zero, it would appear that a useful extension with this method is not possible.

## Summary of Procedure

The steps for record ~~extension~~ augmentation of the mean and variance at the short-record site and constructing a record extension with  $n_e$  observations to EMA are as follows:-

1. Select a hydrologically ~~-relevant~~ relevant longer record site nearby to extend the short-record site of interest. The cross-correlation between the two sites is critical and should be as large as possible. A time series plot of the two series may reveal if the transfer of information using the two sites is reasonable. Does the longer record appear to provide a longer view of what hydrologic events were observed in the shorter record?
2. Investigate the statistical properties and regression relationship between the short and long record sites using base-10 logarithms of the flood flows. If the correlation coefficient estimated with equation 8.9 exceeds a critical value ( $\hat{\rho} > 0.80$ ), record extension may be suitable. Otherwise, it may be advisable to

DRAFT: August 31, 2017

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

use the short record with a weighted skew estimate for frequency analysis, or other techniques such as quantile regression for the site of interest. Here the threshold for extension is given in terms of the true cross-correlation between the series. In practice, a hydrologist must use the computed sample correlation or a regional estimator.

3. Estimate the sample statistics for the concurrent  $n_1$  records using equations 8.1–8.6~~and~~–8.6, and then the mean and variance ~~for the complete extended record~~ estimators for the  $y$ -site based upon the complete record available at the long record site ( $n_1 + n_2$ ) using equations 8.7 and 8.8.
4. Estimate the ~~extended total effective~~ record length  $n_1 + n_e$  using equation 8.19;  $n_e$  is the number of observations that need to be added to the  $y$ -series.
5. Estimate the ~~regression model parameters (equations~~ extension parameters (eqs. 8.23 and 8.24). Use the model (~~equation eq. 8.20~~) to ~~estimate the generate the additional~~  $n_e$  flow values to extend the record for the short site, with the most recent  $n_e$  observations from the  $n_2$  non-concurrent record. Check to make sure the extension is reasonable. As an example, if a short extension of 10 years included the 1 or 2 largest floods in the much longer  $x$ -series, that short extension may misrepresent the likely skewness coefficient for the short-record site. In such cases the  $n_e$   $x$ -observations used to generate the extension might be selected to be a sequence of years that generated a more typical skewness coefficient (considering the different  $n_e$  extensions that could be made) for the extended  $n_1 + n_e$  record.
6. A frequency analysis can then be performed using this extended record flow series  $n_1 + n_e$ .

This procedure was recently developed to focus on  $n_e$  years. Additional work may be needed to select the  $n_e$  years. As noted above and in Matalas and Jacobs (1964), the basic model assumption of joint normality with a skewness equal to zero between the two sites needs to be reviewed.

~~An example of applying MOVE is shown below~~An example is provided in the next section. Supplemental material to perform these calculations and this example is available at <https://acwi.gov/hydrology/Frequency/b17c/>.

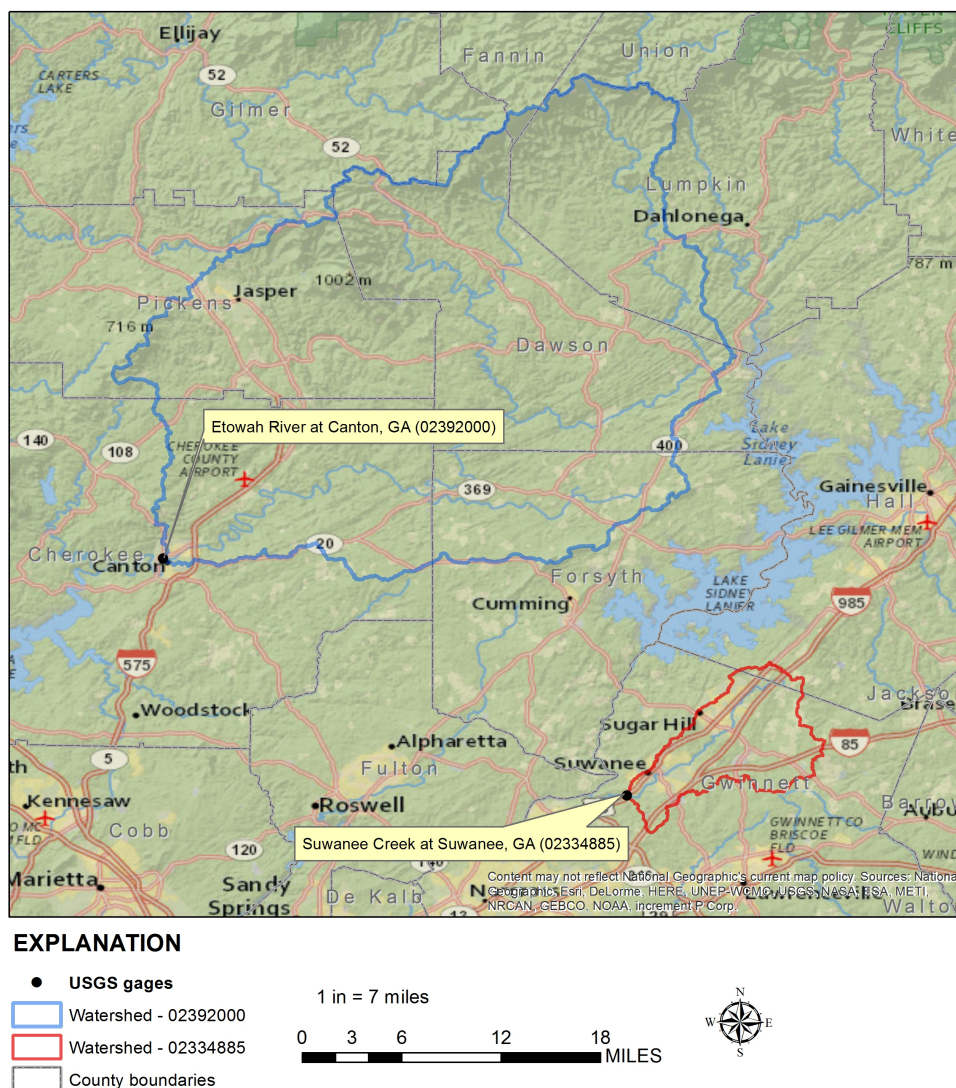
---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017



**Figure 8.1.** Map showing location of Suwanee Creek at Suwanee, Georgia. U.S. Geological Survey station 02334885.

## MOVE Example — Suwanee Creek at Suwanee, Georgia

An example of record extension using MOVE is given for Suwanee Creek at Suwanee, Georgia (station 02334885) where the drainage area is 47.0 square miles. The watershed is located in north-central Georgia as shown in figure 8.1. There are 20 years of record at station 02334885 from 1985 to 2004, which is relatively short, as shown in figure 8.2. The analysis of the 20 years of record for Suwanee Creek provided low estimates of the flood discharges such as the 0.01 AEP flood compared to other long-term stations in the region. There is a nearby gaging station on the Etowah River at Canton, Georgia (station 02392000), which has 113 years of record from 1892 through 2004 and a drainage area of 613 square miles. The annual peak data for the Etowah River were used to extend the record for Suwanee Creek by 13 additional years to obtain more reasonable flood estimates such as the 0.01 AEP flood.

The concurrent annual peak data available at the Etowah River and Suwanee Creek through 2004 are given in table 8.1. For 8 of the 20 years of record, the annual maximum peak flow occurred on the same flood event

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN. IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS). IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

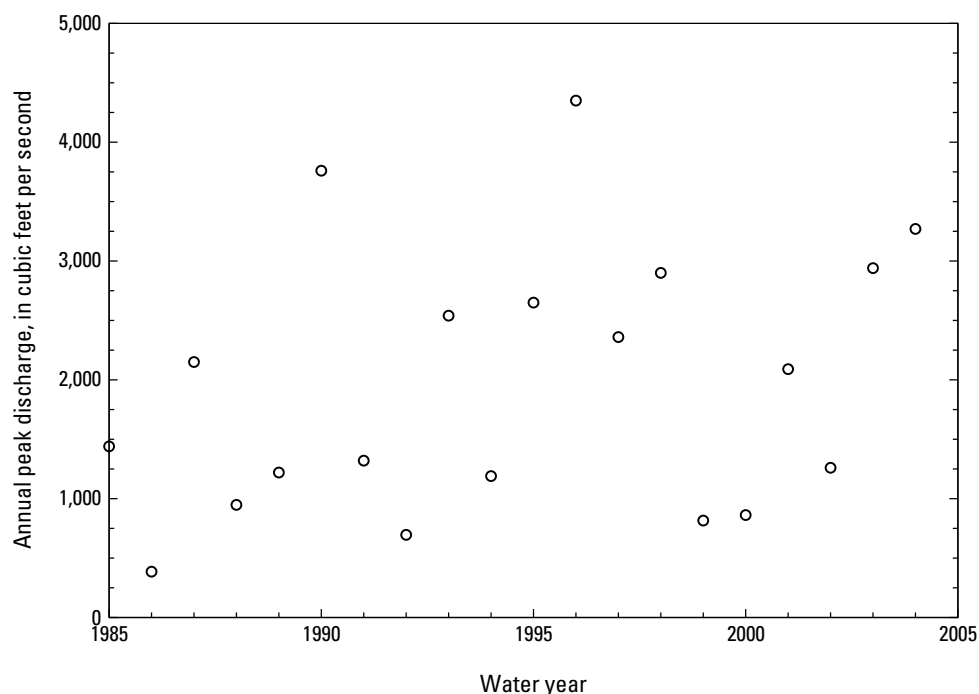
DRAFT: August 31, 2017

for Suwanee Creek and the Etowah River. However, for 12 of the 20 years of concurrent record, the annual peak flows corresponded to different flood events. For the purposes of record extension, concurrent flood peaks are those that occurred in the same water year, not on the same flood event.

The annual peak flows for the Etowah River at Canton, Georgia, the long record station, are plotted in figure 8.3 for 1892 to 2004. As shown in figure 8.3, there were several large floods that were recorded on the Etowah River prior to 1985 when systematic data collection began at Suwanee Creek. The period of systematic record at Suwanee Creek for 1985 to 2004 does not include several large floods that occurred in 1892, 1916, and 1919 at the nearby Etowah River gaging station. These large floods, as well as information provided by other events in the 1892–1984 period, can be used through record augmentation to improve the mean and variance and extend the flood record at Suwanee Creek.

The 20 years of concurrent record for Suwanee Creek and the Etowah River are plotted in figure 8.4 on a log-log scale. As shown in figure 8.4, the logarithms of the annual peak flows define a linear relation with a  $R^2$  value of 0.7258. The correlation coefficient is 0.8519, which is higher than the critical value ( $\hat{\rho} > 0.80$ ) for both the mean and variance. This suggests that the mean and standard deviation from the extended record will be improved by use of a longer record. Even though the Etowah River is much larger than Suwanee Creek, there is a strong correlation in annual peak flows that facilitates record extension. The linear relation shown in figure 8.4 is the ordinary least squares regression line computed using the logarithms of the data.

The 20 years of record at Suwanee Creek from 1985 to 2004 provides low estimates of the flood discharges, like the 0.01 annual exceedance probability discharge because major floods that occurred prior to systematic data collection are not considered in the frequency analysis. The period 1985 to 2004 was a relatively dry period as compared to the period prior to 1985, as can be observed from the long-term Etowah River record shown in figure 8.3.



**Figure 8.2.** Graph showing annual peak flows for Suwanee Creek, U.S. Geological Survey station 02334885, the short record station, from 1985 to 2004.

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017



## 118 Guidelines for Determining Flood Flow Frequency – Bulletin 17C

The flood records for Suwanee Creek were extended using record augmentation of the mean and variance and MOVE, based on the annual peak-flow data for the Etowah River from 1892 to 1984 and overlapping period from 1985-2004. As shown in figure 8.4, there is a strong linear relation between the logarithms of the annual peak flows. The estimated correlation coefficient for the overlapping 1985-2004 period is 0.8519. The sample statistics for the gages are

$y_i$  = logarithmic discharge for Suwanee Creek for year  $i$ ,

$x_i$  = logarithmic discharge for Etowah River for year  $i$ ,

$n_1$  = concurrent or overlapping period of record, 20 years (1985-2004),

$n_2$  = non-concurrent period of record, 93 years (1892-1984),

$\bar{y}_1$  = logarithmic mean for Suwanee Creek for concurrent period = 3.215 log units,

$\bar{x}_1$  = logarithmic mean for Etowah River for concurrent period = 3.984 log units,

$\bar{x}_2$  = logarithmic mean for Etowah River for non-concurrent period = 4.079 log units,

$s_{y1}$  = logarithmic standard deviation for Suwanee Creek for concurrent period = 0.279 log units,

$s_{x1}$  = logarithmic standard deviation for Etowah River for concurrent period = 0.214 log units, and

$s_{x2}$  = logarithmic standard deviation for Etowah River for non-concurrent period = 0.219 log units.

Using equation 8.19, the estimated extended record length  $n_e$  is 13 years.

The extended  $n_e$  years of record  $y_i$  for Suwanee Creek were estimated using the most recent non-overlapping 13 years from 1972 through 1984. Using the Matalas-Jacobs equations 8.7 – 8.11, the augmented mean and variance are  $\hat{\mu}_y = 3.304$  and  $\hat{\sigma}_y^2 = 0.081$ . Solving equations 8.20 – 8.24, the MOVE equation in logarithmic

**Table 8.1.** Summary of concurrent observed annual peak data for the Etowah River and Suwanee Creek from 1985 to 2004.

Water year	Etowah River annual peak streamflow (cfs)	Suwanee Creek Annual Peak Streamflow (cfs)
1985	5,030	1,440
1986	3,090	386
1987	12,200	2,150
1988	9,340	948
1989	9,080	1,220
1990	27,100	3,760
1991	5,940	1,320
1992	7,660	696
1993	10,900	2,540
1994	9,420	1,190
1995	10,500	2,650
1996	19,500	4,350
1997	11,300	2,360
1998	15,000	2,900
1999	5,530	816
2000	8,900	862
2001	9,270	2,090
2002	7,100	1,260
2003	13,600	2,940
2004	15,300	3,270

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

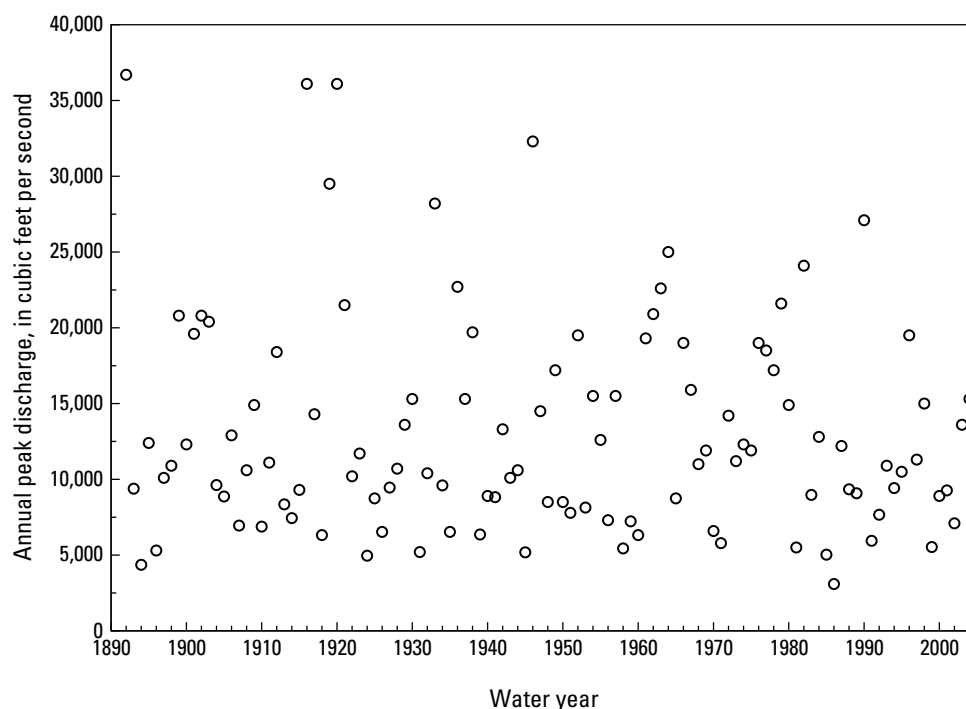
linear form that is used to generate  $n_e$  flows is

$$y_i = 3.436 + 0.8478(x_i - 4.141) \quad (8.25)$$

where  $x_i$  are the base-10 logarithms of the Etowah River flows for the period 1972-1984 and

$$Q_i = 10^{y_i} \quad (8.26)$$

where  $Q_i$  are the extended discharges in cfs for Suwanee Creek. Equation 8.26 was used to estimate annual peak flows for Suwanee Creek for the period 1972 to 1984, thereby extending the record an additional 13 years with data from the Etowah River. Extended flow estimates for Suwanee Creek are listed in table 8.2. Original flow estimates from the long-record site (Etowah River) are listed in table 8.3.



**Figure 8.3.** Graph showing annual peak flows for the Etowah River, U.S. Geological Survey station 02392000, the long record station, from 1892 to 2004.

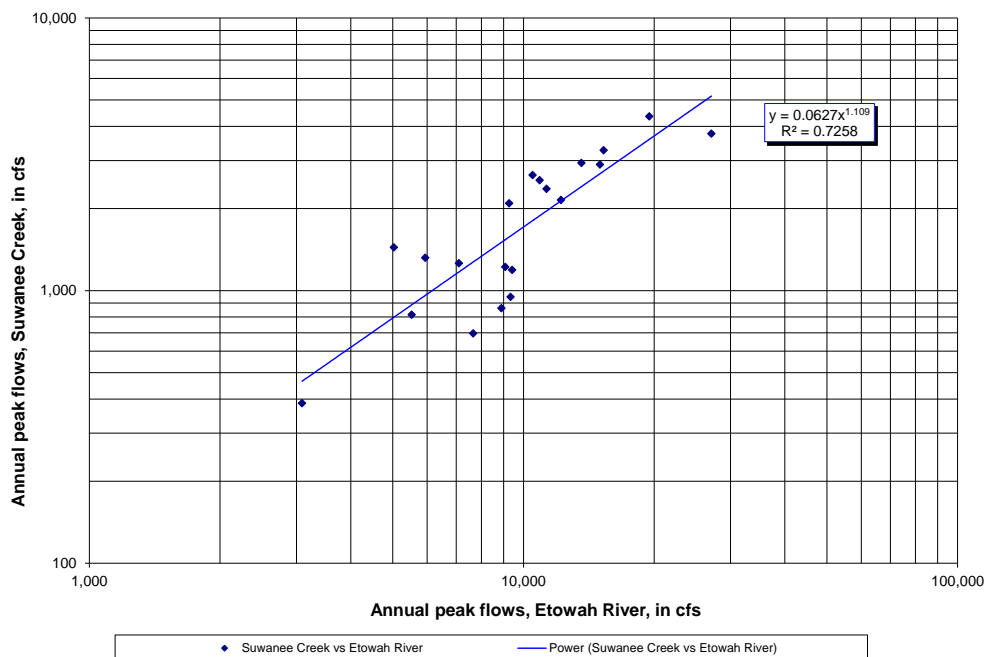
---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017



**Figure 8.4.** Graph of 20 concurrent years of record for Suwanee Creek and the Etowah River for the period 1985 to 2004 with the ordinary least squares regression line (measured in cubic feet per second, cfs).

DRAFT: August 31, 2017

---PROVISIONAL---  
 THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
 PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
 PEER REVIEW PLAN.  
 IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
 GEOLOGICAL SURVEY (USGS).  
 IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
 TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

**Table 8.2.** MOVE extended record for 13 years (1972–1984) for Suwanee Creek at Suwanee, Georgia (station 02334885).

Water year	Annual peak streamflow (cfs)
1972	2,790
1973	2,280
1974	2,470
1975	2,400
1976	3,570
1977	3,490
1978	3,280
1979	3,980
1980	2,910
1981	1,250
1982	4,370
1983	1,890
1984	2,560

DRAFT: August 31, 2017

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

**Table 8.3.** Flood records for 93 years (1892–1984) for the Etowah River at Canton, Georgia (station 02335000).

<b>Water year</b>	<b>Annual peak streamflow (cfs)</b>	<b>Water year</b>	<b>Annual peak streamflow (cfs)</b>	<b>Water year</b>	<b>Annual peak streamflow (cfs)</b>
1892	36,700	1923	11,700	1954	15,500
1893	9,380	1924	4,960	1955	12,600
1894	4,360	1925	8,740	1956	7,300
1895	12,400	1926	6,530	1957	15,500
1896	5,300	1927	9,460	1958	5,440
1897	10,100	1928	10,700	1959	7,230
1898	10,900	1929	13,600	1960	6,320
1899	20,800	1930	15,300	1961	19,300
1900	12,300	1931	5,200	1962	20,900
1901	19,600	1932	10,400	1963	22,600
1902	20,800	1933	28,200	1964	25,000
1903	20,400	1934	9,600	1965	8,740
1904	9,620	1935	6,530	1966	19,000
1905	8,870	1936	22,700	1967	15,900
1906	12,900	1937	15,300	1968	11,000
1907	6,950	1938	19,700	1969	11,900
1908	10,600	1939	6,360	1970	6,590
1909	14,900	1940	8,900	1971	5,790
1910	6,880	1941	8,820	1972	14,200
1911	11,100	1942	13,300	1973	11,200
1912	18,400	1943	10,100	1974	12,300
1913	8,350	1944	10,600	1975	11,900
1914	7,440	1945	5,180	1976	19,000
1915	9,300	1946	32,300	1977	18,500
1916	36,100	1947	14,500	1978	17,200
1917	14,300	1948	8,500	1979	21,600
1918	6,320	1949	17,200	1980	14,900
1919	29,500	1950	8,500	1981	5,500
1920	36,100	1951	7,790	1982	24,100
1921	21,500	1952	19,500	1983	8,970
1922	10,200	1953	8,140	1984	12,800

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

## Appendix 9. Weighting of Independent Estimates

---

1

2

DRAFT: August 31, 2017

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.



DRAFT: August 31, 2017

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

## Weighting of Independent Estimates

The uncertainty of ~~peak-flow~~peak-flow statistics, such as the one-percent annual exceedance probability (AEP) flow at a streamgage (or site), can be reduced by combining the at-site estimate with an independent regional estimate to obtain a weighted estimate of the flow statistic at the site. The analysis assumes that the two estimators are independent, unbiased, and that their estimates of the variances are reliable and consistent. A common use of this approach is to combine at-site flood frequency analysis estimates of flood quantiles with flood quantile estimates obtained by regional regression. In that case, methods developed by ~~federal~~Federal agencies allow computation of weighted estimates using this method. In other cases, independent flood quantile estimates might be available based upon precipitation estimates with rainfall-runoff models. Alternative weighting procedures are evaluated by Griffis and Stedinger (2007a).

The following procedure is suggested for adjusting flow frequency estimates based upon short records (typically less than 20 years) to reflect flood experience in nearby hydrologically similar watersheds, as described in the ~~Section~~section Comparisons of Frequency Curves. The method is based upon the assumption that the estimates are independent, which for practical purposes is true in most situations. If the at-site estimate is used to develop the quantile regression model, quantile weighting is inappropriate as the estimates are not independent. Prior to weighting and combining estimates, the quantiles and variances of the estimates need to be evaluated.

## Weighting Method

As stated in the ~~Section~~section Flood Distribution, the Pearson Type III distribution with log transformation of the ~~peak-flow~~peak-flow data should be the base method for the analysis of annual series data. Thus, the ~~peak-flow~~peak-flow statistic  $Q_i$  (such as the 0.01 AEP) is transformed using base 10 logarithms:

$$X_i = \log_{10} Q_i \quad (9.1)$$

where  $Q_i$  is the estimated ~~peak-flow~~peak-flow statistic at site  $i$ , and  $X_i$  is the log-transformed variable. All subsequent operations are performed on the transformed variable  $X_i$ . The weighted estimate is calculated using variances as ÷:

$$X_{weighted,i} = \frac{X_{site,i}V_{reg,i} + X_{reg,i}V_{site,i}}{V_{site,i} + V_{reg,i}} \quad (9.2)$$

where all  $X$  and  $V$  variables are in  $\log_{10}$  units,  $X_{weighted,i}$  is the weighted estimate for site  $i$ ,  $X_{site,i}$  is the at-site estimate at site  $i$ ,  $X_{reg,i}$  is the regional estimate at site  $i$ ,  $V_{site,i}$  is the variance of the at-site estimate at site  $i$ , and  $V_{reg,i}$  is the variance of the regional estimate at site  $i$ .

As described in ~~Appendix~~appendix 7, the Expected Moments Algorithm (EMA) provides a direct fit of the log-Pearson Type III distribution, which includes an estimate of the variance  $V_{site,i}$  corresponding to each computed AEP.

For independent  $X_{site,i}$  and  $X_{reg,i}$ , the variance of the weighted estimate for site  $i$  is calculated (with all  $V$  variables in  $\log_{10}$  units) as ÷:

$$V_{weighted,i} = \frac{V_{site,i}V_{reg,i}}{V_{site,i} + V_{reg,i}} \quad (9.3)$$

Confidence intervals on the weighted estimated can also be calculated. For example, upper and lower 95%

DRAFT: August 31, 2017

PROVISIONAL  
THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.  
IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).  
IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

## 126 Guidelines for Determining Flood Flow Frequency – Bulletin 17C

confidence limits on the weighted quantile estimate are calculated as :-

$$95\%_{-}CI_i = \left[ 10^{(X_{weighted,i} - 1.96\sqrt{V_{weighted,i}})}, 10^{(X_{weighted,i} + 1.96\sqrt{V_{weighted,i}})} \right], \quad (9.4)$$

1 and note that  $X_{weighted,i}$ ,  $V_{weighted,i}$ , and  $CI_i$  must be calculated separately for each site  $i$  for each AEP of interest.

### 2 Example

3 A flood frequency analysis at a basin (site  $i$ ) using the EMA produces an estimate of 861 ft<sup>3</sup>/s for the 0.01  
4 AEP with a log space variance  $V_{site} = 0.0281$ . Based on hydrologically similar nearby basins, an independent  
5 regional estimate of the 0.01 AEP is 718 ft<sup>3</sup>/s with a log space variance  $V_{reg} = 0.085$ . By substituting these  
6 values into the above equations, the following weighted estimates are obtained.

Using equation 9.1, the log transformed flow values are computed as :-

$$X_{site} = \log_{10}(861) = 2.94$$

$$X_{reg} = \log_{10}(718) = 2.86.$$

Using equation 9.2, the weighted log transformed flow is computed as :-

$$X_{weighted} = \frac{2.94 * 0.085 + 2.86 * 0.028}{0.028 + 0.085} = 2.92$$

and the peak flow  $Q_{weighted}$  is

$$Q_{weighted} = 10^{2.92} = 832 \text{ ft}^3/\text{s}.$$

Using equation 9.3, the variance of the weighted log transformed flow is computed as:

$$V_{weighted} = \frac{0.028 * 0.085}{0.028 + 0.085} = 0.021.$$

Using equation 9.4, a ~~95%~~95-percent confidence interval on the weighted estimate is computed as

$$\begin{aligned} 95\%_{-}CI_i &= \left[ 10^{(2.92 - 1.96\sqrt{0.021})}, 10^{(2.92 + 1.96\sqrt{0.021})} \right] \\ &= [432 \text{ ft}^3/\text{s}, 1,600 \text{ ft}^3/\text{s}]. \end{aligned}$$

DRAFT: August 31, 2017

---PROVISIONAL---  
THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.  
IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).  
IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

## Appendix 10. Examples

---

1

2

DRAFT: August 31, 2017

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

## Flood Frequency Steps and Examples

The main steps to perform a flood frequency analysis using these Guidelines and some representative flood frequency examples are presented in this appendix. The main emphasis is on the data, flow intervals, and threshold inputs to EMA to illustrate applications. Steps to perform a flood frequency analysis at a gaged site are as follows:

1. Define the purpose of the flood frequency study and quantiles of interest (e.g., for example,  $Q_{50}$ ,  $Q_{100}$ ).
2. Obtain peak-flow data at the gaged site, including any historical, paleoflood, and botanical data, regional skew estimates, and other precipitation, climate, and regional information as described in [Appendix appendix 3](#). Perform field studies to collect additional extreme flood, historical, and paleoflood data as needed for the quantiles of interest.
3. Estimate perception thresholds and flow intervals with guidance in [Appendix appendix 3](#).
4. Conduct exploratory data analysis as described in [Appendix appendix 4](#) and evaluate data.
5. For short-record sites ( $n < 20$ ), consider performing record extension as described in [Appendix appendix 8](#).
6. Estimate flood frequency with EMA using the estimated at-site record and regional skew with recommended software listed at <https://acwi.gov/hydrology/Frequency/b17c/>.
7. Consider performing quantile weighting described in [Appendix appendix 9](#), in particular for short-record sites ( $n < 20$ ) or to combine estimates from rainfall-runoff models, depending on the study purpose and quantiles of interest.

The following flood frequency examples illustrate application of the techniques recommended in this guide. Annual flood peak data for seven stations have been selected to illustrate fitting the log-Pearson Type III distribution when one or more of the following are present in a peak flood record at a gage site:

1. Systematic record;
2. Potentially Influential Low Floods (PILFs);
3. Broken record;
4. Historical data;
5. ~~Crest Stage~~ [Crest-stage](#) gage censored data;
6. Historic data and PILFs; and
7. Paleoflood data.

The gaging stations and types of data used in each example are listed in [Table table 10.1](#). The ~~U.S. Geological Survey~~ [USGS](#) *PeakFQ* program (~~Section section~~ Software and Examples) is used for most of the examples shown here; *PeakfqSA* is used for the historical and paleoflood examples. Input and output files from ~~U.S. Geological Survey~~ [USGS](#) *PeakFQ* software used to create the examples, as well as example files for the U.S. Army Corps of Engineers HEC-SSP software, are available on the HFAWG web page at <https://acwi.gov/hydrology/Frequency/b17c/>.

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

## 130 Guidelines for Determining Flood Flow Frequency – Bulletin 17C

These examples are meant to illustrate the main concepts presented in these Guidelines. They are not meant to be ~~all-inclusive~~all inclusive, and are to be used for example purposes only. The examples are provided with sufficient documentation to ensure that results are reproducible based on the input data shown. Given a single input ~~data-set~~dataset, two users will obtain the same answer. Different answers by users may be possible with different interpretations of the data and inputs. The input and output results shown are not intended to be used for making ~~floodplain~~flood-plain-management decisions at specific locations.

It is important to note that ~~for~~ for the purposes of flood frequency analysis, water years are used in these examples to define the years in which annual peak flows occur. A water year is defined as the ~~12-month~~12-month period from October 1 to September 30. The water year is designated by the calendar year in which it ends and which includes 9 of the 12 months.

Weighted skew was only used in ~~Example~~example 1: Systematic data; it was not used in ~~Examples~~examples 2-7. In order to clearly illustrate how the EMA and MGBT screening for PILFs are used in flood frequency analysis, only the at-site skew at each station was used.

**Table 10.1.** Summary of flood frequency examples.

Example Type <del>No.</del> <u>No.</u>	USGS Station <del>No.</del> <u>Station no.</u>	Station Name <del>name</del> <u>name</u>	Systematic	Historical	Broken	Censored	PILF
1	Systematic	01134500 Moose River at Victory, <del>VT</del> <u>Vt.</u>	X				
2	PILF	11274500 Orestimba Creek near Newman, <del>CA</del> <u>Calif.</u>	X				X
3	Broken record	01614000 Back Creek near Jones Springs, <del>WV</del> <u>W. Va.</u>	X		X		
4	Historical data	07099500 Arkansas River at Pueblo, <del>CO</del> <u>Colo.</u>	X	X	X	X	
5	Crest <del>Stage</del> <u>stage</u> (censored)	05489490 Bear Creek at Ottumwa, <del>IA</del> <u>Iowa</u>	X			X	
6	Historic data + PILFs	09480000 Santa Cruz River at Lochiel, <del>AZ</del> <u>Ariz.</u>	X	X			X
7	Paleoflood data	11446500 American River at Fair Oaks, <del>CA</del> <u>Calif.</u>	X	X	X	X	

---PROVISIONAL---

DRAFT: August 31, 2017

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.  
IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).  
IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.



## Systematic Record Example—Moose River at Victory, Vermont

This example illustrates the use of EMA and the MGBT to perform a flood frequency analysis on a gage site with a record comprised of systematic annual flood peaks.

For this example, USGS gage 01134500 Moose River at Victory, Vermont, is used. The Moose River is located in the northeastern part of the State and flows mostly from north to south through very hilly terrain. The Moose River basin is approximately 75 square miles of nearly all forest (Olson, 2014). Historically, it was an important logging area and some logging still continues today. Attempts at farming in the basin have generally failed due to the presence of shallow rocky soil. There are a small number of villages in the basin, but overall it is sparsely populated with only a few miles of paved roadway. There is also a large bog approximately one third of a mile upstream from the gage. The bog is part of the 5,000-acre Victory Basin Wildlife management area. While there is no streamflow regulation in the basin, the bog attenuates peaks in the basin.

USGS gage 01134500 has an annual peak record consisting of 68 peaks, beginning in 1947 and ending in 2014. The annual peaks are listed in table 10.2 and can be downloaded from USGS NWIS at [http://nwis.waterdata.usgs.gov/nwis/peak/?site\\_no=01134500&agency\\_cd=USGS&](http://nwis.waterdata.usgs.gov/nwis/peak/?site_no=01134500&agency_cd=USGS&).

## EMA Representation of Peak Flow Data for Flood Frequency Analysis

As described in the Data Representation using Flow Intervals and Perception Thresholds section, when using EMA the annual peak flow for every water year during the historical period is described by a flow interval ( $Q_{Y,lower}$ ,  $Q_{Y,upper}$ ) for each water year  $Y$ . For peaks whose values are known and are not censored, the flow interval can be described as ( $Q_{Y,lower} = Q_Y$ ,  $Q_{Y,upper} = Q_Y$ ). For example, as shown in table 10.2, the peak for the 1947 water year is recorded as 2,080 cfs. This peak is known and is not censored, thus the flow interval for the 1947 water year is ( $Q_{1947,lower} = 2,080$ ,  $Q_{1947,upper} = 2,080$ ). In this example, the flow values are known for all the years where the gage was in operation. Table 10.3 contains the EMA flow intervals for each water year in the record for gage 01134500.

As described in the Data Representation using Flow Intervals and Perception Thresholds section, EMA distinguishes among sampling properties by employing perception thresholds denoted ( $T_{Y,lower}$ ,  $T_{Y,upper}$ ) for each year  $Y$ , which reflect the range of flows that would have been measured or recorded had they occurred. Perception thresholds describe the range of measurable potential discharges and are independent of the actual peak discharges that have occurred. The lower bound,  $T_{Y,lower}$ , represents the smallest peak flow that would result in a recorded flow in water year  $Y$ . For most peaks at most gages,  $T_{Y,upper}$  is assumed to be infinite, as bigger floods that might exceed the measurement capability of the streamgage are determined through study of high-water marks and other physical evidence of the flood. For periods of continuous, full-range peak-flow record, the perception threshold is represented by ( $T_{Y,lower} = 0$ ,  $T_{Y,upper} = \infty$ ), where  $T_{Y,lower} = 0$  is the gage-base discharge. Table 10.4 contains the EMA perception thresholds for each water year in the record for USGS gage 01134500.

The annual peaks, as well as their corresponding EMA flow intervals and perception thresholds, can be displayed graphically. Figure 10.1 contains a graphical representation of the recorded annual peaks, EMA flow intervals, and EMA perception thresholds. This graph of the data is simple for USGS gage 01134500, as each year in the record has a recorded peak and the perception threshold for the entire period of record spans from ( $T_{Y,lower} = 0$ ,  $T_{Y,upper} = \infty$ ), thus indicating that all peaks were able to be recorded.

## Results from Flood Frequency Analysis

A flood frequency analysis at USGS gage 01134500 was performed using the EMA flow intervals and perception thresholds as shown in table 10.3 and table 10.4. The output from an at-site flood frequency analysis

DRAFT: August 31, 2017

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.  
IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).  
IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

**Table 10.2.** U.S. Geological Survey gage 01134500 Moose River at Victory, Vermont annual peak-flow record consisting of 68 peaks from 1947 to 2014. This table contains the date of the annual peak recorded at the gage, the water year of the annual peak, and the corresponding annual peak in cubic feet per second (cfs).

Date of peak streamflow	Water year	Annual peak streamflow (cfs)	Date of peak streamflow	Water year	Annual peak streamflow (cfs)	Date of peak streamflow	Water year	Annual peak streamflow (cfs)
1947-04-13	1947	2,080	1970-04-25	1970	3,010	1993-04-17	1993	1,900
1948-03-28	1948	1,670	1971-05-04	1971	1,490	1994-04-17	1994	2,760
1949-03-28	1949	1,480	1972-05-05	1972	2,920	1995-08-06	1995	4,536
1950-04-21	1950	2,940	1973-07-01	1973	4,940	1996-04-24	1996	2,160
1950-12-05	1951	1,560	1973-12-22	1974	2,550	1996-12-02	1997	1,860
1952-06-02	1952	2,380	1975-04-20	1975	1,250	1998-03-31	1998	2,680
1953-03-27	1953	2,720	1976-04-02	1976	2,670	1999-09-18	1999	1,540
1954-04-23	1954	2,860	1977-03-31	1977	2,020	2000-05-11	2000	2,110
1955-04-15	1955	2,620	1978-05-10	1978	1,460	2001-04-25	2001	2,950
1956-04-30	1956	1,710	1979-03-26	1979	1,620	2002-04-14	2002	2,410
1957-04-22	1957	1,370	1980-04-10	1980	1,460	2003-03-30	2003	2,230
1957-12-21	1958	2,180	1981-02-21	1981	1,570	2003-10-28	2004	1,980
1959-04-04	1959	1,160	1982-04-18	1982	2,890	2005-04-04	2005	1,610
1959-11-29	1960	2,780	1983-05-04	1983	1,840	2005-10-17	2006	2,640
1961-04-24	1961	1,580	1984-05-31	1984	2,950	2007-04-24	2007	1,930
1962-04-08	1962	2,110	1985-04-17	1985	1,380	2008-04-20	2008	1,940
1963-04-22	1963	2,160	1986-03-31	1986	2,350	2009-04-04	2009	1,810
1964-04-15	1964	2,750	1987-03-31	1987	4,180	2010-03-24	2010	1,900
1965-06-14	1965	1,190	1988-04-06	1988	1,700	2010-10-01	2011	3,140
1966-03-26	1966	1,560	1989-04-06	1989	2,200	2012-03-20	2012	1,370
1967-04-03	1967	1,800	1990-03-18	1990	3,430	2013-04-20	2013	2,180
1968-03-24	1968	1,600	1990-12-24	1991	2,270	2014-04-16	2014	4,250
1969-04-29	1969	2,400	1992-04-23	1992	2,180			

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

**Table 10.3.** U.S. Geological Survey gage 01134500 Moose River at Victory, Vermont Expected Moments Algorithm flow intervals for the systematic period from 1947 to 2014. This table contains the water year of the annual peak and the corresponding flow interval defined by lower bound,  $Q_{Y,lower}$ , and upper bound,  $Q_{Y,upper}$ , in cubic feet per second for each water year  $Y$ .

Water Year	$Q_{Y,lower}$	$Q_{Y,upper}$	Comments	Water Year	$Q_{Y,lower}$	$Q_{Y,upper}$	Comments
1947	2,080	2,080		1981	1,570	1,570	
1948	1,670	1,670		1982	2,890	2,890	
1949	1,480	1,480		1983	1,840	1,840	
1950	2,940	2,940		1984	2,950	2,950	
1951	1,560	1,560		1985	1,380	1,380	
1952	2,380	2,380		1986	2,350	2,350	
1953	2,720	2,720		1987	4,180	4,180	
1954	2,860	2,860		1988	1,700	1,700	
1955	2,620	2,620		1989	2,200	2,200	
1956	1,710	1,710		1990	3,430	3,430	
1957	1,370	1,370		1991	2,270	2,270	
1958	2,180	2,180		1992	2,180	2,180	
1959	1,160	1,160		1993	1,900	1,900	
1960	2,780	2,780		1994	2,760	2,760	
1961	1,580	1,580		1995	4,536	4,536	
1962	2,110	2,110		1996	2,160	2,160	
1963	2,160	2,160		1997	1,860	1,860	
1964	2,750	2,750		1998	2,680	2,680	
1965	1,190	1,190		1999	1,540	1,540	
1966	1,560	1,560		2000	2,110	2,110	
1967	1,800	1,800		2001	2,950	2,950	
1968	1,600	1,600		2002	2,410	2,410	
1969	2,400	2,400		2003	2,230	2,230	
1970	3,010	3,010		2004	1,980	1,980	
1971	1,490	1,490		2005	1,610	1,610	
1972	2,920	2,920		2006	2,640	2,640	
1973	4,940	4,940		2007	1,930	1,930	
1974	2,550	2,550		2008	1,940	1,940	
1975	1,250	1,250		2009	1,810	1,810	
1976	2,670	2,670		2010	1,900	1,900	
1977	2,020	2,020		2011	3,140	3,140	
1978	1,460	1,460		2012	1,370	1,370	
1979	1,620	1,620		2013	2,180	2,180	
1980	1,460	1,460		2014	4,250	4,250	

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

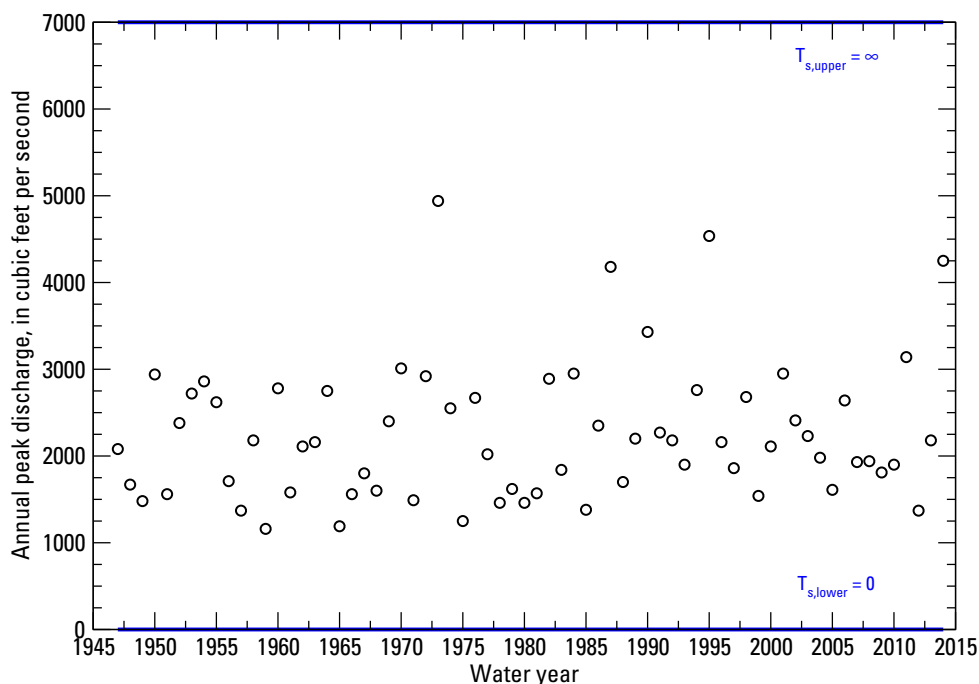
DRAFT: August 31, 2017

**Table 10.4.** U.S. Geological Survey gage 01134500 Moose River at Victory, Vermont Expected Moments Algorithm perception thresholds for the systematic period from 1947 to 2014. This table contains the water year ranges to which each perception threshold applies;  $T_{Y,lower}$ , the lower bound of the perception threshold (in cubic feet per second) for water year  $Y$ ;  $T_{Y,upper}$ , the upper bound of the perception threshold in cfs for water year  $Y$ ; and a comment describing the threshold.

EMA perception threshold				
Start year	End year	$T_{Y,lower}$	$T_{Y,upper}$	Comments
1947	2014	0	Infinity	Continuous systematic record.

using EMA with the Multiple Grubbs-Beck test to screen for potentially influential low floods (PILFs) is shown below. Note that for the analysis described below, weighted skew was used. As described in the section Estimating Regional Skew, an improved estimate of skew can be computed by weighting the station skew with a regional skew (see the section Weighted Skew Coefficient Estimator for details). The regional skew used for this station is 0.44 with a corresponding standard error of 0.28 (MSE=0.078) (Olson, 2014). The at-site skew estimate is 0.397 with a MSE=0.10. The estimated peak flow for selected annual exceedance probabilities can be found in table 10.5, while the fitted frequency curve is displayed in figure 10.2. The final estimated moments were 3.3286 (mean), 0.1403 (standard deviation), and 0.421 (weighted skew).

The results of the above analysis were generated using weighted skew. In order to demonstrate the potential impact of the weighted skew on a flood frequency analysis, here we present the results for the same analysis using solely the station skew and compare the results to those previously obtained using the weighted skew. As shown in figure 10.3, the confidence intervals when using only the station skew are wider for the smaller exceedance probabilities as compared to those in figure 10.2 when the weighted skew is used. It is important to note that this is just one example of the effect of weighted skew on a flood frequency analysis. The impact could be more significant or less significant than shown above, depending on the peak-flow data at the station,



**Figure 10.1.** Graph showing U.S. Geological Survey gage 01134500 Moose River at Victory, Vermont annual peak-flow time series consisting of 68 peaks from 1947 to 2014. Open circles represent recorded systematic peaks.

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

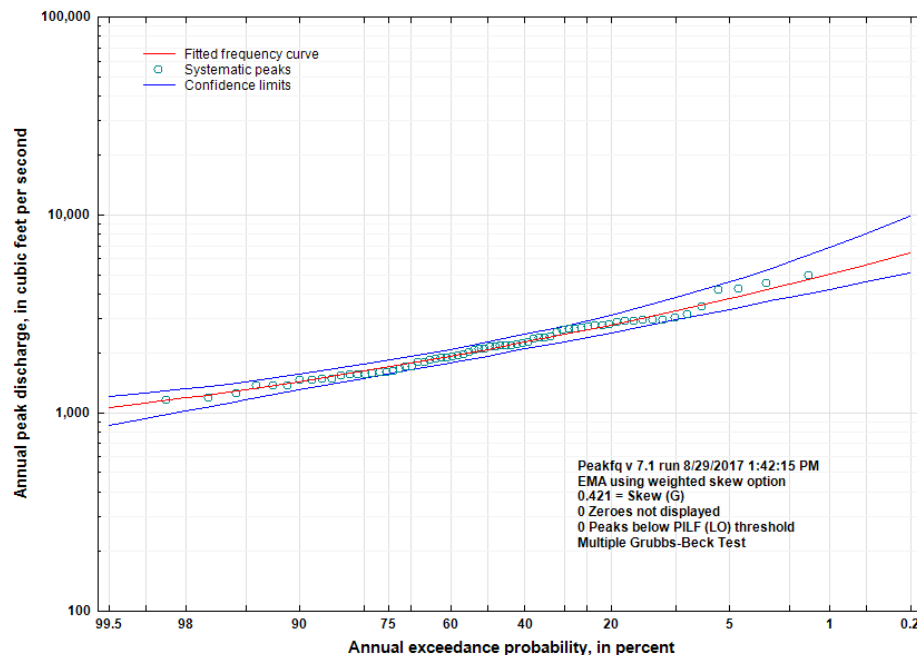
DRAFT: August 31, 2017

**Table 10.5.** Peak-flow quantiles in cubic feet per second for U.S. Geological Survey gage 01134500 based on flood frequency analysis using Expected Moments Algorithm with Multiple Grubbs-Beck Test; variance of estimate shown in log space.

Annual exceedance probability	EMA estimate	Variance of estimate	Lower 2.5% confidence limit	Upper 97.5% confidence limit
0.1	3,261	0.0007	2,933	3,826
0.04	3,911	0.0014	3,426	5,038
0.02	4,422	0.0021	3,782	6,234
0.01	4,957	0.0031	4,131	7,735
0.005	5,519	0.0043	4,476	9,616
0.002	6,313	0.0064	4,929	12,850

as well as the value of the station's corresponding regional skew and the accuracy of that regional skew. 1

The fitted frequency curve computed using EMA with MGBT is displayed in red in figure 10.2. Because the annual peak-flow record contains only systematic peaks with no historic information, no censored peaks, and no PILFs identified by the MGBT, the fitted frequency curve using these flood frequency Guidelines is the same as that from Bulletin 17B. 2  
3  
4  
5



**Figure 10.2.** Graph showing annual exceedance probability plot for U.S. Geological Survey gage 01134500 Moose River at Victory, Vermont based on flood frequency analysis using Expected Moments Algorithm with Multiple Grubbs-Beck Test and weighted skew. The red line is the fitted log-Pearson Type III frequency curve, the blue lines are the upper and lower confidence limits, and the green circles are the systematic peaks.

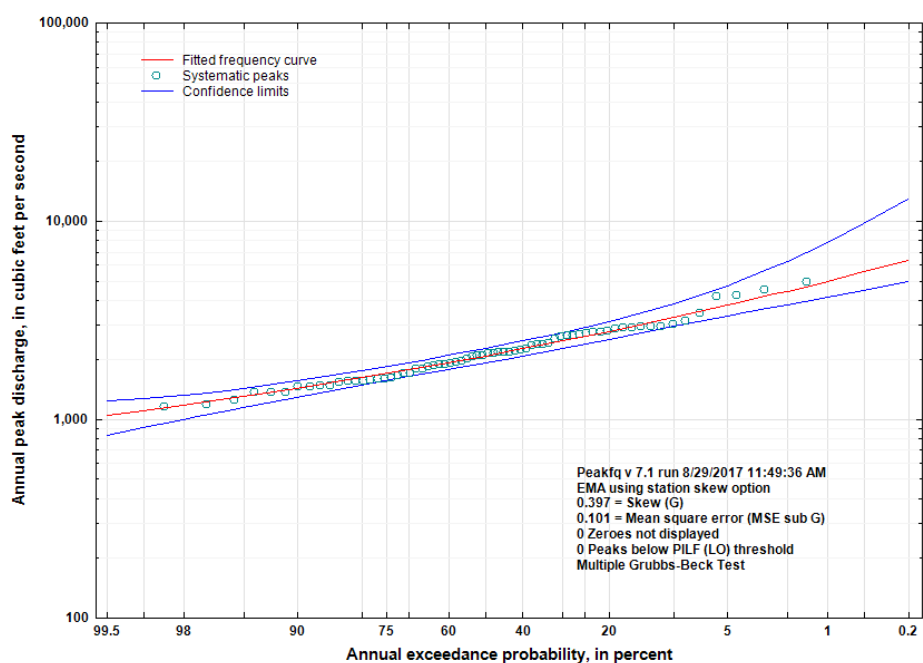
---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017



**Figure 10.3.** Graph showing annual exceedance probability plot for U.S. Geological Survey gage 01134500 Moose River at Victory, Vermont based on flood frequency analysis using Expected Moments Algorithm with Multiple Grubbs-Beck Test and station skew only. The red line is the fitted log-Pearson Type III frequency curve, the blue lines are the upper and lower bounds of the confidence limits, and the green circles are the systematic peaks.

DRAFT: August 31, 2017

---PROVISIONAL---  
 THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
 PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
 PEER REVIEW PLAN.  
 IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
 GEOLOGICAL SURVEY (USGS).  
 IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
 TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

## PILF Example—Orestimba Creek near Newman, California

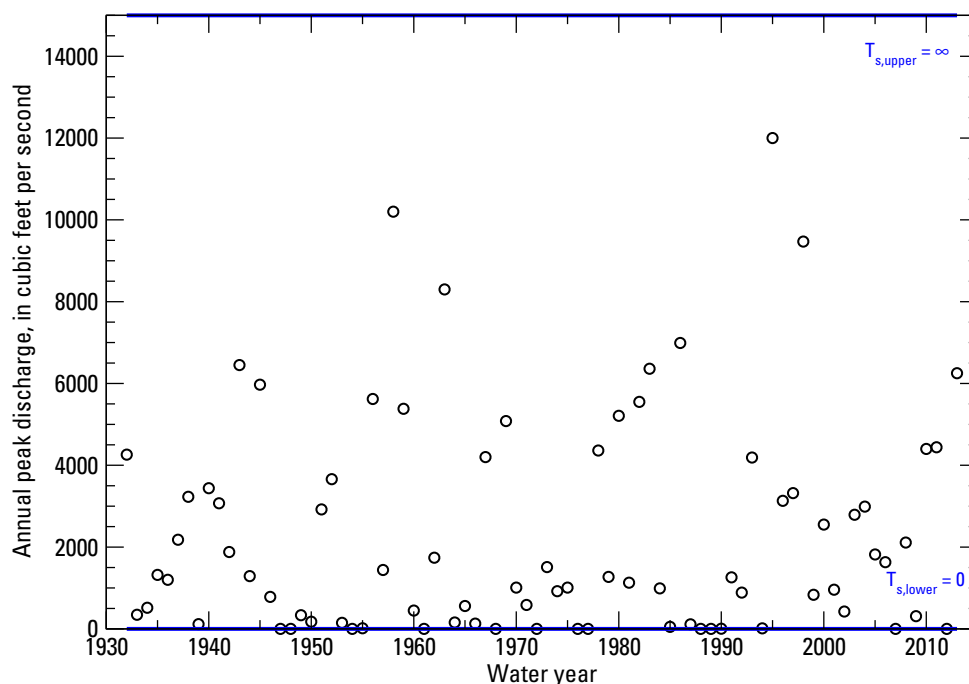
This example demonstrates how the Expected Moments Algorithm (EMA) and the Multiple Grubbs-Beck test (MGBT) can be used to perform a flood frequency analysis on a gage site with a record comprised of systematic annual flood peaks when potentially influential low floods (PILFs) are present.

For this example, USGS gage 11274500 Orestimba Creek near Newman, California, is used (Parrett and others, 2011; Gotvald and others, 2012). Orestimba Creek is a tributary to the San Joaquin River, whose 134-square-mile drainage area lies on the eastern slope of the Diablo Range section of the Coast Range Mountains of California (U.S. Army Corps of Engineers, 2008). The drainage basin has an average basin elevation of 1,551 feet with peak flows usually occurring in late winter. Orestimba Creek is one of the few tributaries in the area to maintain a definite stream channel from the foothills to the San Joaquin River (U.S. Army Corps of Engineers, 2008). Some additional details about this gage are in Gotvald and others (2012).

USGS gage 11274500 has an annual peak record consisting of 82 peaks beginning in 1932 and ending in 2013. The annual peaks are listed in table 10.6 (downloaded from USGS NWIS: [http://nwis.waterdata.usgs.gov/nwis/peak/?site\\_no=11274500&agency\\_cd=USGS&](http://nwis.waterdata.usgs.gov/nwis/peak/?site_no=11274500&agency_cd=USGS&)) and shown in figure 10.4. Of the 82 annual peaks, there are 12 years for which the annual peak is 0 ft<sup>3</sup>/s.

## EMA Representation of Peak Flow Data for Flood Frequency Analysis

As described in the Data Representation using Flow Intervals and Perception Thresholds section, when using EMA the annual peak flow for every water year during the historical period is described by a flow interval ( $Q_{Y,lower}$ ,  $Q_{Y,upper}$ ) for each water year  $Y$ . For peaks whose values are known and are not censored, the flow interval can be described as ( $Q_{Y,lower} = Q_Y$ ,  $Q_{Y,upper} = Q_Y$ ). For example, as shown in table 10.6, the peak for



**Figure 10.4.** Graph showing U.S. Geological Survey gage 11274500 Orestimba Creek near Newman, California annual peak-flow time series consisting of 82 peaks from 1932 to 2013. Open circles represent recorded systematic peaks.

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017



**Table 10.6.** U.S. Geological Survey gage 11274500 Orestimba Creek near Newman, California annual peak-flow record consisting of 82 peaks from 1932 to 2013. This table contains the date of the annual peak recorded at the gage, the water year of the annual peak, and the corresponding annual peak in cubic feet per second (cfs).

Date of peak streamflow	Water year	Annual peak streamflow (cfs)	Date of peak streamflow	Water year	Annual peak streamflow (cfs)	Date of peak streamflow	Water year	Annual peak streamflow (cfs)
1932-02-08	1932	4,260	1960-02-10	1960	448	1988-00-00	1988	0
1933-01-29	1933	345	1961-00-00	1961	0	1989-00-00	1989	0
1934-01-01	1934	516	1962-02-15	1962	1,740	1990-05-28	1990	4
1935-04-08	1935	1,320	1963-02-01	1963	8,300	1991-03-24	1991	1,260
1936-02-13	1936	1,200	1964-01-22	1964	156	1992-02-15	1992	888
1937-02-13	1937	2,180	1965-01-06	1965	560	1993-01-13	1993	4,190
1938-02-11	1938	3,230	1965-12-30	1966	128	1994-02-20	1994	12
1939-03-09	1939	115	1967-01-24	1967	4,200	1995-03-10	1995	12,000
1940-02-27	1940	3,440	1968-00-00	1968	0	1996-02-19	1996	3,130
1941-04-04	1941	3,070	1969-01-25	1969	5,080	1997-01-23	1997	3,320
1942-01-24	1942	1,880	1970-03-01	1970	1,010	1998-02-03	1998	9,470
1943-01-21	1943	6,450	1970-12-21	1971	584	1999-02-09	1999	833
1944-02-29	1944	1,290	1972-00-00	1972	0	2000-02-14	2000	2,550
1945-02-02	1945	5,970	1973-02-11	1973	1,510	2001-03-05	2001	958
1945-12-25	1946	782	1974-03-03	1974	922	2002-01-03	2002	425
1947-00-00	1947	0	1975-03-08	1975	1,010	2002-12-16	2003	2,790
1948-00-00	1948	0	1976-00-00	1976	0	2004-02-25	2004	2,990
1949-03-12	1949	335	1977-00-00	1977	0	2005-02-16	2005	1,820
1950-02-05	1950	175	1978-01-17	1978	4,360	2006-01-02	2006	1,630
1950-12-03	1951	2,920	1979-02-21	1979	1,270	2007-00-00	2007	0
1952-01-12	1952	3,660	1980-02-16	1980	5,210	2008-01-25	2008	2,110
1952-12-07	1953	147	1981-01-29	1981	1,130	2009-02-17	2009	310
1954-00-00	1954	0	1982-01-05	1982	5,550	2010-01-20	2010	4,400
1955-01-19	1955	16	1983-01-24	1983	6,360	2011-03-24	2011	4,440
1955-12-23	1956	5,620	1983-12-25	1984	991	2012-00-00	2012	0
1957-02-24	1957	1,440	1985-02-09	1985	50	2012-12-24	2013	6,250
1958-04-02	1958	10,200	1986-02-19	1986	6,990			
1959-02-16	1959	5,380	1987-03-06	1987	112			

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

the 1932 water year is recorded as 4260 ft<sup>3</sup>/s. This peak is known and is not censored, thus the flow interval for the 1932 water year is ( $Q_{1932,lower} = 4260$ ,  $Q_{1932,upper} = 4260$ ). Table 10.7 contains the EMA flow intervals for each water year in the record for gage 11274500.

As described in the Data Representation using Flow Intervals and Perception Thresholds section, EMA distinguishes among sampling properties by employing perception thresholds denoted ( $T_{Y,lower}$ ,  $T_{Y,upper}$ ) for each year  $Y$ , which reflect the range of flows that would have been measured or recorded had they occurred. Perception thresholds describe the range of measurable potential discharges and are independent of the actual peak discharges that have occurred. The lower bound,  $T_{Y,lower}$ , represents the smallest peak flow that would result in a recorded flow in water year  $Y$ . For most peaks at most gages,  $T_{Y,upper}$  is assumed to be infinite, as bigger floods that might exceed the measurement capability of the streamgage are determined through study of high-water marks and other physical evidence of the flood. For periods of continuous, full-range peak-flow record, the perception threshold is represented by ( $T_{Y,lower} = 0$ ,  $T_{Y,upper} = \infty$ ), where  $T_{Y,lower} = 0$  is the gage-base discharge. Table 10.8 contains the EMA perception thresholds for each water year in the record for USGS gage 11274500.

## Results from Flood Frequency Analysis

A flood frequency analysis at USGS gage 11274500 was performed using the EMA flow intervals and perception thresholds as shown in table 10.7 and table 10.8. The output from an at-site flood frequency analysis using EMA with the Multiple Grubbs-Beck test to screen for PILFs is shown below. Note that station skew was used, thus allowing the focus to be on the at-site data. The fitted frequency curve is displayed in figure 10.5 with estimates provided in table 10.9. The final estimated moments were 3.0227 (mean), 0.6821 (standard deviation), and -0.929 (station skew).

As shown in figure 10.5, the PILF threshold  $T_{PILF}$  established by the MGBT is 782 ft<sup>3</sup>/s, with a significance level equal to 0.0007. Thus, all 30 annual peaks (including 12 zeros) less than 782 ft<sup>3</sup>/s are censored and re-coded in the framework of EMA with flow intervals of ( $Q_{Y,lower} = 0$ ,  $Q_{Y,upper} = 782$ ). The MGBT threshold also has the effect of adjusting the lower bound of the perception threshold. Thus for the entire historical period from 1932 to 2013, the perception threshold based on  $T_{PILF}$  is ( $T_{Y,lower} = 782$ ,  $T_{Y,upper} = \infty$ ). As shown in figure 10.5, by censoring the 30 smallest peaks in the record, the smallest annual exceedance probability peaks are well fit by the frequency curve (red line).

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

## 140 Guidelines for Determining Flood Flow Frequency – Bulletin 17C

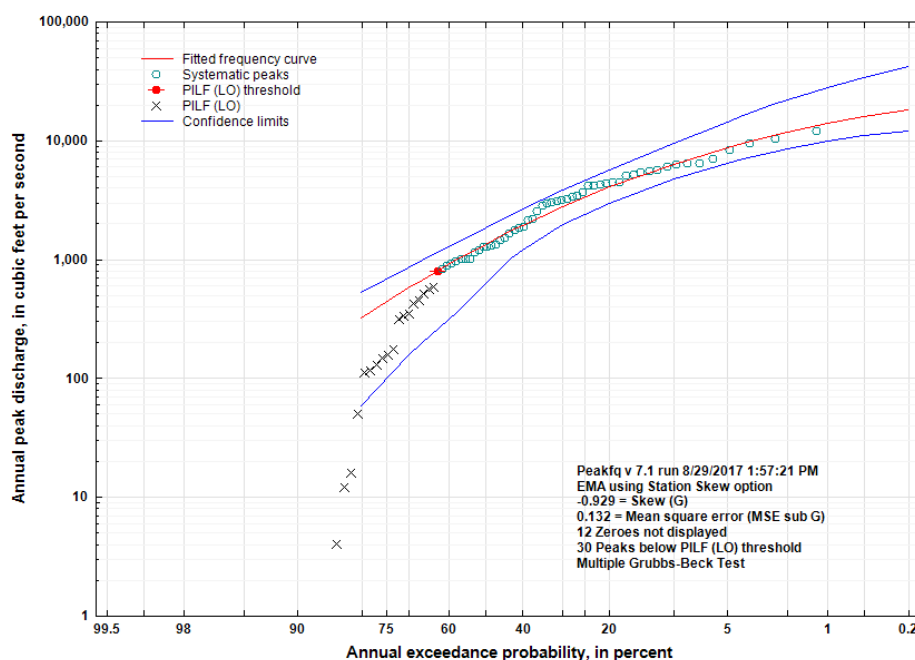
**Table 10.7.** U.S. Geological Survey gage 11274500 Orestimba Creek near Newman, California Expected Moments Algorithm flow intervals for the systematic period from 1932 to 2013. This table contains the water year of the annual peak and the corresponding flow interval defined by lower bound,  $Q_{Y,lower}$ , and upper bound,  $Q_{Y,upper}$ , in cubic feet per second (cfs) for each water year  $Y$ .

Water year	$Q_{Y,lower}$	$Q_{Y,upper}$	Comments	Water year	$Q_{Y,lower}$	$Q_{Y,upper}$	Comments
1932	4,260	4,260		1973	1,510	1,510	
1933	345	345		1974	922	922	
1934	516	516		1975	1,010	1,010	
1935	1,320	1,320		1976	0	0	Zero flow.
1936	1,200	1,200		1977	0	0	Zero flow.
1937	2,180	2,180		1978	4,360	4,360	
1938	3,230	3,230		1979	1,270	1,270	
1939	115	115		1980	5,210	5,210	
1940	3,440	3,440		1981	1,130	1,130	
1941	3,070	3,070		1982	5,550	5,550	
1942	1,880	1,880		1983	6,360	6,360	
1943	6,450	6,450		1984	991	991	
1944	1,290	1,290		1985	50	50	
1945	5,970	5,970		1986	6,990	6,990	
1946	782	782		1987	112	112	
1947	0	0	Zero flow.	1988	0	0	Zero flow.
1948	0	0	Zero flow.	1989	0	0	Zero flow.
1949	335	335		1990	4	4	
1950	175	175		1991	1,260	1,260	
1951	2,920	2,920		1992	888	888	
1952	3,660	3,660		1993	4,190	4,190	
1953	147	147		1994	12	12	
1954	0	0	Zero flow.	1995	12,000	12,000	
1955	16	16		1996	3,130	3,130	
1956	5,620	5,620		1997	3,320	3,320	
1957	1,440	1,440		1998	9,470	9,470	
1958	10,200	10,200		1999	833	833	
1959	5,380	5,380		2000	2,550	2,550	
1960	448	448		2001	958	958	
1961	0	0	Zero flow.	2002	425	425	
1962	1,740	1,740		2003	2,790	2,790	
1963	8,300	8,300		2004	2,990	2,990	
1964	156	156		2005	1,820	1,820	
1965	560	560		2006	1,630	1,630	
1966	128	128		2007	0	0	Zero flow.
1967	4,200	4,200		2008	2,110	2,110	
1968	0	0	Zero flow.	2009	310	310	---PROVISIONAL---
1969	5,080	5,080		2010	4,400	4,400	THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.
1970	1,010	1,010		2011	4,440	4,440	IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).
1971	584	584		2012	0	0	Zero flow.
1972	0	0	Zero flow.	2013	6,250	6,250	THIS DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

**Table 10.8.** U.S. Geological Survey gage 11274500 Orestimba Creek near Newman, California Expected Moments Algorithm perception thresholds for the period from 1932 to 2013. This table contains the water year ranges to which each perception threshold applies;  $T_{Y,lower}$  the lower bound of the perception threshold (in cfs) for water year  $Y$ ,  $T_{Y,upper}$  the upper bound of the perception threshold in cfs for water year  $Y$ ; and a comment describing the threshold.

EMA perception threshold				
Start year	End year	$T_{Y,lower}$	$T_{Y,upper}$	Comments
1932	2013	0	Infinity	Continuous systematic record.



**Figure 10.5.** Graph showing annual exceedance probability plot for USGS Gage 11274500 Orestimba Creek near Newman, California based on flood frequency analysis using Expected Moments Algorithm with Multiple Grubbs-Beck Test. The red line is the fitted log-Pearson Type III frequency curve, the blue lines are the upper and lower confidence limits, the green circles are the systematic peaks, the solid red circle with a line through it is the PILF threshold as identified by the MGBT, and the black x's are the PILFs identified by the MGBT.

DRAFT: August 31, 2017

---PROVISIONAL---  
THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.  
IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).  
IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

**Table 10.9.** Peak-flow quantiles in cubic feet per second for USGS Gage 11274500 Orestimba Creek near Newman, California based on flood frequency analysis using Expected Moments Algorithm with Multiple Grubbs-Beck Test; variance of estimate shown in log space.

Annual exceedance probability	EMA estimate	Variance of estimate	Lower 2.5% confidence limit	Upper 97.5% confidence limit
0.5	1,339	0.0078	620.6	1,840
0.2	4,026	0.0045	2,965	5,595
0.1	6,328	0.0049	4,686	9,394
0.04	9,426	0.0061	6,944	16,110
0.02	11,690	0.0073	8,489	21,920
0.01	13,820	0.0088	9,813	27,730
0.005	15,800	0.0106	10,910	33,500
0.002	18,150	0.0135	12,040	41,610

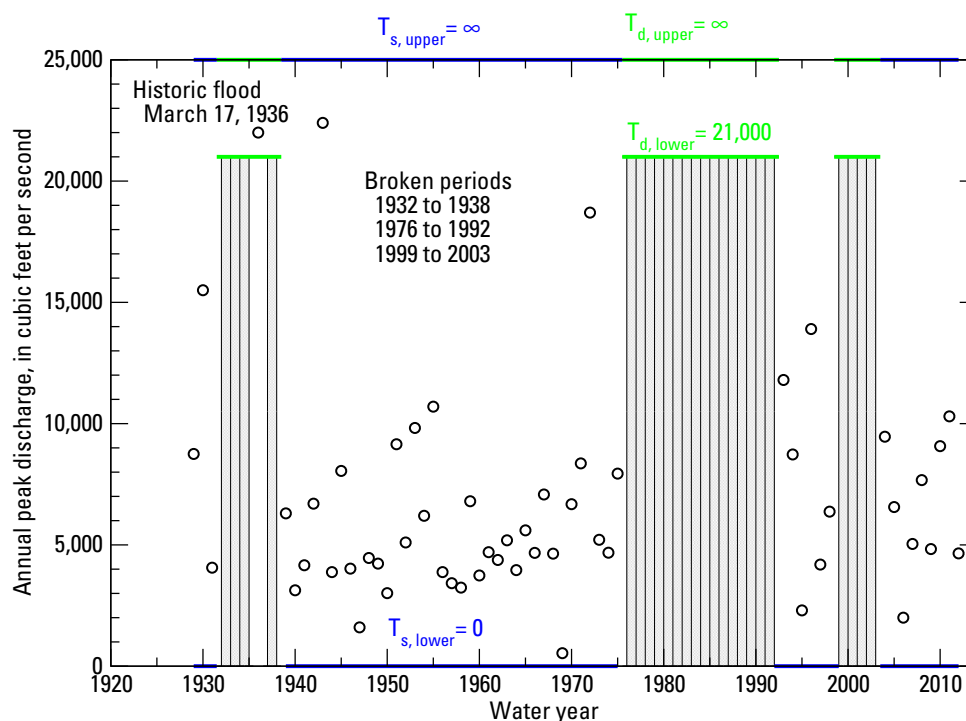
DRAFT: August 31, 2017

---PROVISIONAL---  
 THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
 PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
 PEER REVIEW PLAN.  
 IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
 GEOLOGICAL SURVEY (USGS).  
 IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
 TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

## Broken Record Example—Back Creek near Jones Springs, West Virginia

This example illustrates the use of EMA for a broken record, as described in the section Broken, Incomplete, and Discontinued Records. For this example, USGS gage 01614000 Back Creek near Jones Springs, West Virginia, is used. Back Creek is a tributary to the Potomac River; the 235-square-mile watershed lies within the Valley and Ridge province in West Virginia (Wiley and Atkins, 2010).

USGS gage 01614000 has an annual peak record consisting of 56 peaks, beginning in 1929 and ending in 2012. There are three “broken record” periods where the gage was discontinued: 1932–1937, 1976–1991, and 1999–2003. Thus, there are 28 years of missing data at this gage during the period 1929–2012. There is a historic flood that occurred outside the period of gaging record on March 17, 1936. This flood is noted in the USGS Annual Water Data Report for this gage, available in the peak-flow file, and there is historical information available for this large flood (Grover, 1937). The annual peaks are listed in table 10.10 and shown in figure 10.6. Of the 56 annual peaks, the October 1942 flood slightly exceeds the March 1936 historic flood peak. Based on the historical flood information in Grover (1937) for the 1936 flood, and the large regional floods and historical floods described by Wiley and Atkins (2010) in West Virginia for the period 1888–1996, information from the March 1936 flood is used as a perception threshold to represent the 28 years of missing information.



**Figure 10.6.** Graph showing U.S. Geological Survey gage 01614000 Back Creek near Jones Springs, West Virginia annual peak-flow time series consisting of 56 peaks from 1929 to 2012. Flood intervals are shown as black vertical bars with caps that represent lower and upper flow estimates. The grey shaded areas represents floods of unknown magnitude less than the perception threshold  $T_{d,lower}$  during the broken record periods. The green lines represent the range in which floods would have been measured or recorded for the broken record periods 1932–1938, 1976–1992, and 1999–2003, with lower and upper perception thresholds  $T_{d,lower}$  (21,000 cfs) and  $T_{d,upper}$  estimated from the March 1936 historic flood. The perceptible range for the systematic (gage) periods  $T_{s,lower}$ ,  $T_{s,upper}$  (0,  $\infty$ ) is shown as blue lines.

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

## EMA Representation of Peak Flow Data for Flood Frequency Analysis

As described in the Data Representation using Flow Intervals and Perception Thresholds section, when using EMA, the annual peak flow for every water year during the historical period is described by a flow interval ( $Q_{Y,lower}$ ,  $Q_{Y,upper}$ ) for each water year  $Y$ . For peaks whose values are known and are not censored, the flow interval can be described as ( $Q_{Y,lower} = Q_Y$ ,  $Q_{Y,upper} = Q_Y$ ). In this example, the flow values are known for all the years where the gage was in operation. Table 10.11 contains the EMA flow intervals for each water year in the record for gage 01614000. Missing years are described by perception thresholds.

As described in the Data Representation using Flow Intervals and Perception Thresholds section, EMA distinguishes among sampling properties by employing perception thresholds denoted ( $T_{Y,lower}$ ,  $T_{Y,upper}$ ) for each year  $Y$ , which reflect the range of flows that would have been measured or recorded had they occurred. Perception thresholds describe the range of measurable potential discharges and are independent of the actual peak discharges that have occurred. The lower bound,  $T_{Y,lower}$ , represents the smallest peak flow that would result in a recorded flow in water year  $Y$ . For most peaks at most gages,  $T_{Y,upper}$  is assumed to be infinite, as bigger floods that might exceed the measurement capability of the streamgage are determined through study of high-water marks and other physical evidence of the flood. For periods of continuous, full-range peak-flow record, the perception threshold is represented by ( $T_{Y,lower} = 0$ ,  $T_{Y,upper} = \infty$ ), where  $T_{Y,lower} = 0$  is the gage-base discharge. In this example, there are missing years that are described by the 1936 historical flood magnitude. Based on the March 1936 large historical flood (Grover, 1937) and the regional historical flood information available for the largest floods in West Virginia (Wiley and Atkins, 2010), it is known that floods at this location would have been estimated (or recorded), had they exceeded approximately 21,000 cfs. Table 10.12 contains the EMA perception thresholds for each water year in the record, including missing periods, for USGS gage 01614000.

## Results from Flood Frequency Analysis

A flood frequency analysis at USGS Gage 01614000 was performed using the EMA flow intervals and perception thresholds as shown in table 10.11 and table 10.12. The output from an at-site flood frequency analysis using EMA with the Multiple Grubbs-Beck test to screen for PILFs is shown below. Note that station skew was used, thus allowing the focus to be on the at-site data. The fitted frequency curve is displayed in figure 10.7 with estimates provided in table 10.13. The final estimated moments were 3.7598 (mean), 0.2434 (standard deviation), and 0.144 (station skew).

As shown in figure 10.7, there are two floods that exceed the historical threshold (21,000 cfs): the March 1936 flood and the October 1942 flood. Using MGBT, two PILFs were identified, with a threshold equal to 2,000 ft<sup>3</sup>/s and a significance level equal to 0.0881. Two annual peaks less than 2,000 ft<sup>3</sup>/s are censored and re-coded in the framework of EMA with flow intervals of ( $Q_{Y,lower} = 0$ ,  $Q_{Y,upper} = 2000$ ). The MGBT threshold also has the effect of adjusting the lower bound of the perception threshold. Thus for the entire historical period from 1929 to 2012, with the exception of the missing years, the perception threshold is ( $T_{Y,lower} = 2000$ ,  $T_{Y,upper} = \infty$ ). For the broken-record years covered by historical information, the lower threshold  $T_{Y,lower} = 21000$  (table 10.12). As shown in figure 10.7, by censoring the one smallest peak in the record, the remaining smallest annual exceedance probability peaks and the largest floods are well fit by the frequency curve (red line).

---PROVISIONAL---

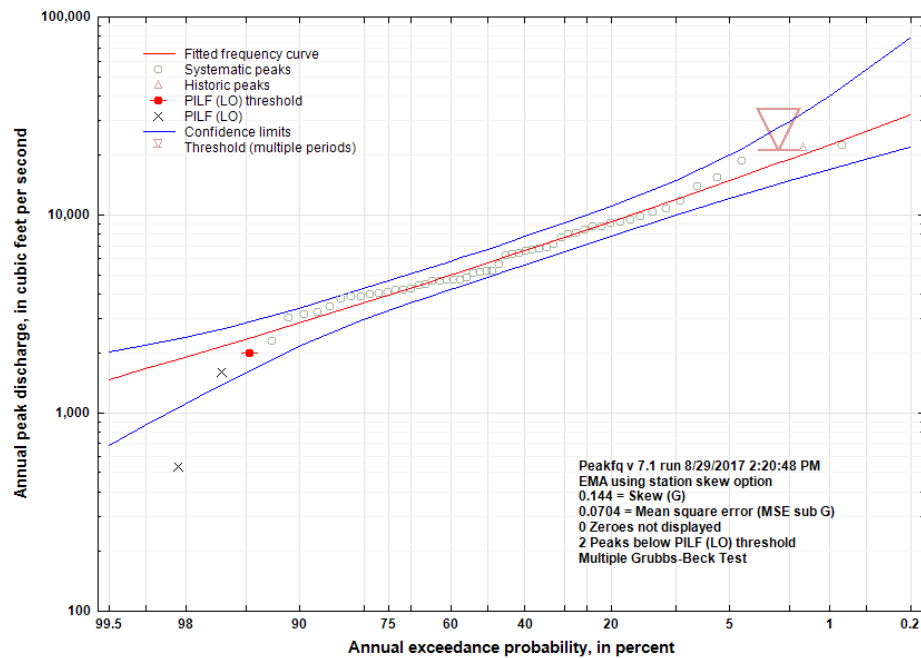
THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017





**Figure 10.7.** Graph showing annual exceedance probability plot for USGS Gage 01614000 Back Creek near Jones Springs, West Virginia based on flood frequency analysis using Expected Moments Algorithm with Multiple Grubbs-Beck Test. The red line is the fitted log-Pearson Type III frequency curve, the blue lines are the upper and lower confidence limits, the green circles are the systematic peaks, the solid red circle with a line through it is the PILF threshold as identified by the MGBT, and the black x's are the PILFs identified by the MGBT. The red triangle with the horizontal line represents the lower limit of the historical perception threshold (21,000 cfs).

DRAFT: August 31, 2017

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

**Table 10.10.** U.S. Geological Survey gage 01614000 Back Creek near Jones Springs, West Virginia annual peak-flow record consisting of 56 peaks from 1929 to 2012, including the 1936 historical flood. This table contains the date of the annual peak recorded at the gage, the water year of the annual peak, and the corresponding annual peak in cubic feet per second (cfs). Horizontal lines indicate broken-record years.

Date of peak streamflow	Water year	Annual peak streamflow (cfs)	Date of peak streamflow	Water year	Annual peak streamflow (cfs)	Date of peak streamflow	Water year	Annual peak streamflow (cfs)
1929-04-17	1929	8,750	1954-03-02	1954	6,200	1972-12-09	1973	5,210
1929-10-23	1930	15,500	1955-08-19	1955	10,700	1973-12-27	1974	4,680
1931-05-08	1931	4,060	1956-03-15	1956	3,880	1975-03-20	1975	7,940
1936-03-17	1936	22,000	1957-02-10	1957	3,420	1993-03-05	1993	11,800
1939-02-04	1939	6,300	1958-03-27	1958	3,240	1994-05-08	1994	8,730
1940-04-20	1940	3,130	1959-06-03	1959	6,800	1995-01-16	1995	2,300
1941-04-06	1941	4,160	1960-05-09	1960	3,740	1996-01-19	1996	13,900
1942-05-22	1942	6,700	1961-02-19	1961	4,700	1996-11-09	1997	4,190
1942-10-15	1943	22,400	1962-03-22	1962	4,380	1998-03-21	1998	6,370
1944-03-24	1944	3,880	1963-03-20	1963	5,190	2004-09-29	2004	9,460
1945-09-18	1945	8,050	1964-01-10	1964	3,960	2005-03-29	2005	6,560
1946-06-03	1946	4,020	1965-03-06	1965	5,600	2005-11-30	2006	2,000
1947-03-15	1947	1,600	1966-09-21	1966	4,670	2007-04-16	2007	5,040
1948-04-14	1948	4,460	1967-03-08	1967	7,080	2008-04-21	2008	7,670
1948-12-31	1949	4,230	1968-03-17	1968	4,640	2009-05-05	2009	4,830
1950-02-02	1950	3,010	1969-02-02	1969	536	2010-03-14	2010	9,070
1950-12-05	1951	9,150	1970-07-10	1970	6,680	2011-04-17	2011	10,300
1952-04-28	1952	5,100	1970-11-13	1971	8,360	2012-03-01	2012	4,650
1952-11-22	1953	9,820	1972-06-22	1972	18,700			

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

**Table 10.11.** U.S. Geological Survey gage 01614000 Back Creek near Jones Springs, West Virginia Expected Moments Algorithm flow intervals for the systematic period from 1929 to 2012. This table contains the water year of the annual peak and the corresponding flow interval defined by lower bound,  $Q_{Y,lower}$ , and upper bound,  $Q_{Y,upper}$ , in cubic feet per second (cfs) for each water year  $Y$ . Horizontal lines indicate broken-record years.

Water year	$Q_{Y,lower}$	$Q_{Y,upper}$	Comments	Water year	$Q_{Y,lower}$	$Q_{Y,upper}$	Comments
1929	8,750	8,750		1963	5,190	5,190	
1930	15,500	15,500		1964	3,960	3,960	
1931	4,060	4,060		1965	5,600	5,600	
1936	22,000	22,000	Historic flood.	1966	4,670	4,670	
1939	6,300	6,300		1967	7,080	7,080	
1940	3,130	3,130		1968	4,640	4,640	
1941	4,160	4,160		1969	536	536	
1942	6,700	6,700		1970	6,680	6,680	
1943	22,400	22,400		1971	8,360	8,360	
1944	3,880	3,880		1972	18,700	18,700	
1945	8,050	8,050		1973	5,210	5,210	
1946	4,020	4,020		1974	4,680	4,680	
1947	1,600	1,600		1975	7,940	7,940	
1948	4,460	4,460		1993	11,800	11,800	
1949	4,230	4,230		1994	8,730	8,730	
1950	3,010	3,010		1995	2,300	2,300	
1951	9,150	9,150		1996	13,900	13,900	
1952	5,100	5,100		1997	4,190	4,190	
1953	9,820	9,820		1998	6,370	6,370	
1954	6,200	6,200		2004	9,460	9,460	
1955	10,700	10,700		2005	6,560	6,560	
1956	3,880	3,880		2006	2,000	2,000	
1957	3,420	3,420		2007	5,040	5,040	
1958	3,240	3,240		2008	7,670	7,670	
1959	6,800	6,800		2009	4,830	4,830	
1960	3,740	3,740		2010	9,070	9,070	
1961	4,700	4,700		2011	10,300	10,300	
1962	4,380	4,380		2012	4,650	4,650	

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

**Table 10.12.** U.S. Geological Survey gage 01614000 Back Creek near Jones Springs, West Virginia Expected Moments Algorithm perception thresholds for the systematic period from 1929 to 2012. This table contains the water year ranges to which each perception threshold applies,  $T_{Y,lower}$  the lower bound of the perception threshold (in cfs) for water year  $Y$ ,  $T_{Y,upper}$  the upper bound of the perception threshold in cfs for water year  $Y$ , and a comment describing the threshold.

Start year	End year	EMA perception threshold		Comments
		$T_{Y,lower}$	$T_{Y,upper}$	
1929	1931	0	Infinity	Continuous systematic record.
1932	1938	21,000	Infinity	Missing record with historical information.
1939	1975	0	Infinity	Continuous systematic record.
1976	1992	21,000	Infinity	Missing record with historical information.
1993	1998	0	Infinity	Continuous systematic record.
1999	2003	21,000	Infinity	Missing record with historical information.
2004	2012	0	Infinity	Continuous systematic record.

**Table 10.13.** Peak-flow quantiles in cubic feet per second for USGS Gage 01614000 Back Creek near Jones Springs, West Virginia based on flood frequency analysis using Expected Moments Algorithm with Multiple Grubbs-Beck Test; variance of estimate shown in log space.

Annual exceedance probability	EMA estimate	Variance of estimate	Lower 2.5% confidence limit	Upper 97.5% confidence limit
0.5	5,675	0.0013	4,819	6,676
0.2	9,179	0.0015	7,745	11,040
0.1	11,890	0.0019	9,879	14,880
0.04	15,770	0.0031	12,690	21,900
0.02	18,980	0.0045	14,820	29,440
0.01	22,480	0.0065	16,950	39,610
0.005	26,290	0.0090	19,060	53,190
0.002	31,860	0.0133	21,840	78,210

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

## Historical Record Example—Arkansas River at Pueblo, Colorado

This example illustrates the use of EMA for a historical record with several large floods (described in the section Historical Flood Information) and paleoflood information. The Arkansas River at Pueblo Dam near Pueblo, Colorado, is presented to illustrate the use of EMA with extensive historical information, paleoflood information, and the ability to place the record June 1921 flood in longer time context. The largest historic floods are described as interval data, and multiple thresholds are needed to effectively extend the discontinued streamgaging record after the dam was built. Paleoflood data are also included for this Reclamation dam safety application. Details of this example are presented in England and others (2006) and England and others (2010). Peak discharge probability estimates were made at four paleoflood sites on the Arkansas River at Pueblo State Park, Parkdale, at Loma Linda and at Adobe Park. We focus on the Pueblo State Park site flood frequency in this example; flood frequency results for other locations, as well as regional frequency results are presented in England and others (2006).

For this example, peak discharge estimates on the Arkansas River at Pueblo State Park are combined from USGS gaging stations at Portland (07097000) (years 1975–6), near Portland (07099200) (1974), and near Pueblo (07099500) (years 1864–1973), and are used with Pueblo reservoir records (years 1977–2004) in order to gain a complete record of all large floods that exceeded approximately 10,000 ft<sup>3</sup>/s for the period of record. Some of these data, particularly the historical flood estimates, were obtained from USGS Water-Supply Papers, Colorado Division of Water Resources Records, historical accounts of the June 1921 flood, and other sources. As such, they do not directly correspond to peak flows in the USGS NWIS database. These gaging stations were previously documented and analyzed by England and others (2006); see also England and others (2010).

The annual peaks are listed in table 10.14 and shown in figure 10.8. Of the 85 annual peaks, including historical information, the June 3, 1921 peak (Follansbee and Jones, 1922; Munn and Savage, 1922) is the largest. The total combined gage record length, excluding historical data, is 110 years (1895–2004) (fig. 10.8). The largest peak discharge estimates from these gages were unaffected by upstream regulation. Reviews of available historical information (Follansbee and Jones, 1922; Munn and Savage, 1922; Follansbee and Sawyer, 1948) indicated there was historical flood information at the site for frequency analysis. The historical record was estimated to begin in 1859, resulting in a 146-year period (1859–2004). Three historical floods were included: June 1864, July 1893, and May 1894. The magnitudes of these floods were large relative to the floods in the gaging record; estimates within a range were based on Follansbee and Sawyer (1948) and included in the flood frequency analysis. These estimates have relatively large uncertainties as compared to the smaller floods in the gage record. A paleohydrologic bound of about 840 years (before water year 2004) was estimated at this site for inclusion in the flood frequency curve. The estimate is based on three soils pits, two radiocarbon ages, and hydraulic modeling of a 7,500-foot reach (England and others, 2006). No estimates of individual paleofloods were made at this site, due to the relatively wide channel geometry and the lack of apparent stratigraphic evidence of large paleofloods during the limited field study (England and others, 2010). Peak discharge, historical flood, and nonexceedance-bound data synthesis for flood frequency show that these historical floods are the largest in the record, and combined with the paleoflood data, result in a substantially longer time series (fig. 10.8).

## EMA Representation of Peak Flow Data for Flood Frequency Analysis

As described in the Data Representation using Flow Intervals and Perception Thresholds section, when using EMA the annual peak flow for every water year during the historical period is described by a flow interval ( $Q_{Y,lower}$ ,  $Q_{Y,upper}$ ) for each water year  $Y$ . For peaks whose values are known and are not censored, the flow interval can be described as ( $Q_{Y,lower} = Q_Y$ ,  $Q_{Y,upper} = Q_Y$ ). In this example, the flow values are known for all the years where the gage was in operation. Table 10.15 contains the EMA flow intervals for each water year

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

**Table 10.14.** U.S. Geological Survey gage 07099500 (and others) Arkansas River annual peak-flow record consisting of 85 peaks from 1864 to 1976. This table contains the water year of the annual peak and the corresponding annual peak in cubic feet per second (cfs).

Water year	Annual peak streamflow (cfs)	Water year	Annual peak streamflow (cfs)	Water year	Annual peak streamflow (cfs)
1864	>41,000	1921	>80,000	1950	8,700
1893	>20,000	1922	8,850	1951	9,300
1894	>35,000	1923	25,600	1952	4,740
1895	6,100	1924	6,510	1953	6,770
1896	16,500	1925	4,930	1954	10,200
1897	4,300	1926	4,520	1955	11,100
1898	7,500	1927	12,400	1956	8,010
1899	8,800	1928	7,800	1957	9,070
1900	7,600	1929	10,500	1958	4,540
1901	11,100	1930	6,050	1959	2,820
1902	30,000	1931	3,560	1960	5,260
1903	10,500	1932	4,380	1961	5,760
1904	8,500	1933	8,630	1962	3,540
1905	8,000	1934	2,580	1963	8,360
1906	11,000	1935	9,880	1964	2,840
1907	6,600	1936	11,200	1965	23,500
1908	7,600	1937	9,300	1966	10,600
1909	5,800	1938	11,200	1967	5,870
1910	8,400	1939	2,910	1968	5,190
1911	3,700	1940	3,860	1969	6,620
1912	10,500	1941	7,560	1970	6,300
1913	7,800	1942	10,300	1971	3,360
1914	7,500	1943	3,320	1972	3,360
1915	17,000	1944	5,980	1973	6,760
1916	8,900	1945	9,290	1974	5,440
1917	6,800	1946	7,050	1975	10,200
1918	9,600	1947	7,280	1976	12,800
1919	6,300	1948	10,900		
1920	8,500	1949	12,800		

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

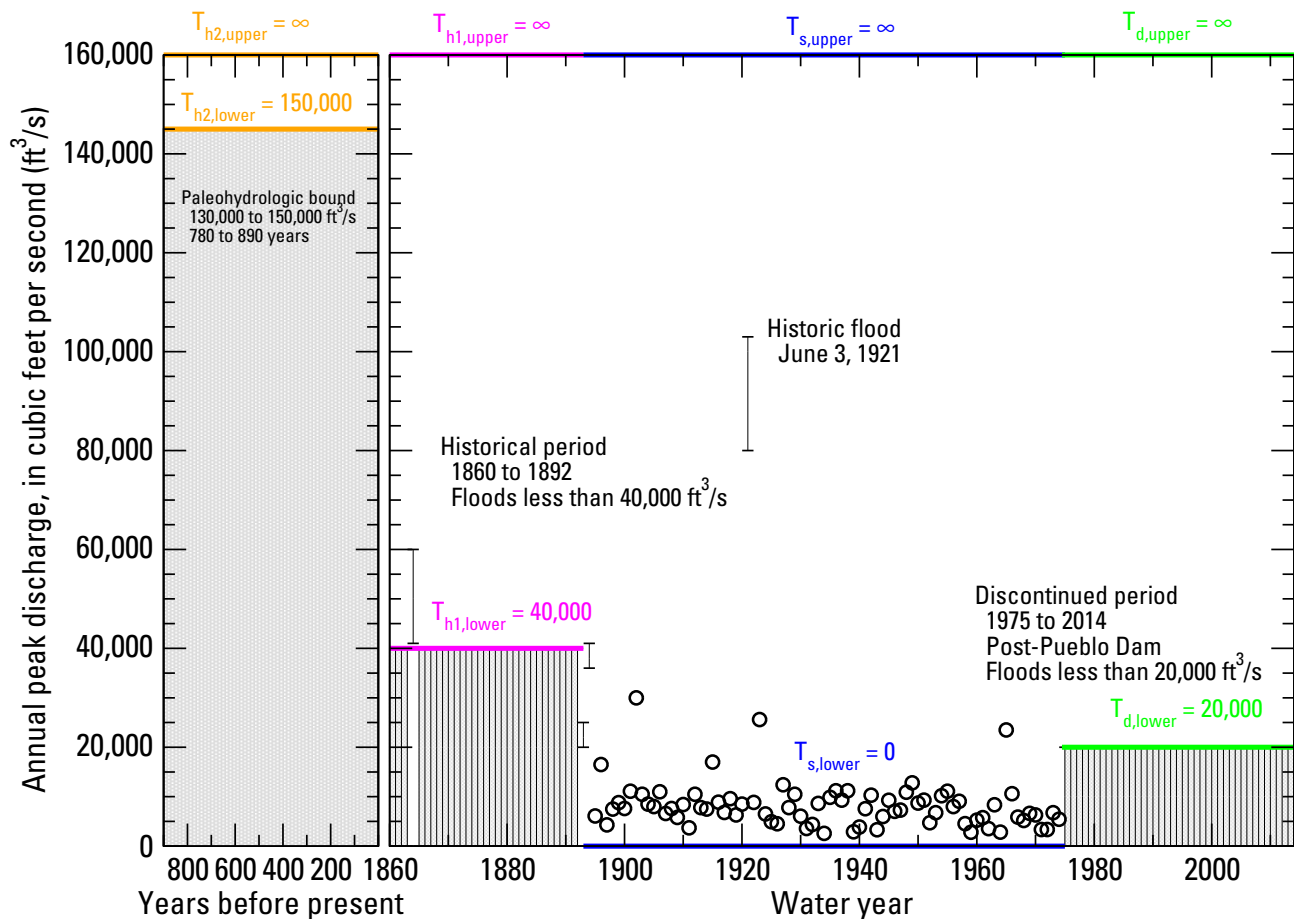
IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

in the record for USGS gage 07099500. The historical period is described by a perception threshold, as is the period after the gage was discontinued (1977–2004).

As described in the Data Representation using Flow Intervals and Perception Thresholds section, EMA distinguishes among sampling properties by employing perception thresholds denoted  $(T_{Y,lower}, T_{Y,upper})$  for each year  $Y$ , which reflect the range of flows that would have been measured or recorded had they occurred. Perception thresholds describe the range of measurable potential discharges and are independent of the actual peak discharges that have occurred. The lower bound,  $T_{Y,lower}$ , represents the smallest peak flow that would result in a recorded flow in water year  $Y$ . For most peaks at most gages,  $T_{Y,upper}$  is assumed to be infinite, as bigger floods that might exceed the measurement capability of the streamgage are determined through study of highwater marks and other physical evidence of the flood. For periods of continuous, full-range peak flow record, the perception threshold is represented by  $(T_{Y,lower} = 0, T_{Y,upper} = \infty)$ , where  $T_{Y,lower} = 0$  is the gage-base discharge. Based on the historical floods and reservoir records, it is known that floods at this location would



**Figure 10.8.** Graph showing peak discharge, historical and paleoflood estimates, Arkansas River at Pueblo State Park. A scale break is used to separate the gage and historical data from the longer paleoflood record. Flood intervals are shown as black vertical bars with caps that represent lower and upper flow estimates, including unobserved estimates in the historical period and historical floods in 1864, 1893, 1894 and 1921. The grey shaded areas represents floods of unknown magnitude less than the perception thresholds for the paleoflood period  $T_{h2,lower}$ , the historical period  $T_{h1,lower}$ , and the discontinued period  $T_{d,lower}$ . Perception threshold ranges are shown as orange lines for the paleoflood period, magenta lines for the historical period, blue lines for the systematic period, and green lines for the discontinued period.

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017



## 152 Guidelines for Determining Flood Flow Frequency – Bulletin 17C

**Table 10.15.** Arkansas River at Pueblo EMA flow intervals for the period from 1864 to 1976.

Water year	Q <sub>Y,lower</sub>	Q <sub>Y,upper</sub>	Comments	Water year	Q <sub>Y,lower</sub>	Q <sub>Y,upper</sub>	Comments
1864	41,000	60,000	Historical flood.	1935	9,880	9,880	
1893	20,000	25,000	Historical flood.	1936	11,200	11,200	
1894	35,000	40,000	Historical flood.	1937	9,300	9,300	
1895	6,100	6,100		1938	11,200	11,200	
1896	16,500	16,500		1939	2,910	2,910	
1897	4,300	4,300		1940	3,860	3,860	
1898	7,500	7,500		1941	7,560	7,560	
1899	8,800	8,800		1942	10,300	10,300	
1900	7,600	7,600		1943	3,320	3,320	
1901	11,100	11,100		1944	5,980	5,980	
1902	30,000	30,000		1945	9,290	9,290	
1903	10,500	10,500		1946	7,050	7,050	
1904	8,500	8,500		1947	7,280	7,280	
1905	8,000	8,000		1948	10,900	10,900	
1906	11,000	11,000		1949	12,800	12,800	
1907	6,600	6,600		1950	8,700	8,700	
1908	7,600	7,600		1951	9,300	9,300	
1909	5,800	5,800		1952	4,740	4,740	
1910	8,400	8,400		1953	6,770	6,770	
1911	3,700	3,700		1954	10,200	10,200	
1912	10,500	10,500		1955	11,100	11,100	
1913	7,800	7,800		1956	8,010	8,010	
1914	7,500	7,500		1957	9,070	9,070	
1915	17,000	17,000		1958	4,540	4,540	
1916	8,900	8,900		1959	2,820	2,820	
1917	6,800	6,800		1960	5,260	5,260	
1918	9,600	9,600		1961	5,760	5,760	
1919	6,300	6,300		1962	3,540	3,540	
1920	8,500	8,500		1963	8,360	8,360	
1921	80,000	103,000	Historical flood.	1964	2,840	2,840	
1922	8,850	8,850		1965	23,500	23,500	
1923	25,600	25,600		1966	10,600	10,600	
1924	6,510	6,510		1967	5,870	5,870	
1925	4,930	4,930		1968	5,190	5,190	
1926	4,520	4,520		1969	6,620	6,620	
1927	12,400	12,400		1970	6,300	6,300	
1928	7,800	7,800		1971	3,360	3,360	
1929	10,500	10,500		1972	3,360	3,360	
1930	6,050	6,050		1973	6,760	6,760	
1931	3,560	3,560		1974	5,440	5,440	---PROVISIONAL---
1932	4,380	4,380		1975	10,200	10,200	THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.
1933	8,630	8,630		1976	12,800	12,800	IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).
1934	2,580	2,580					IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

**Table 10.16.** USGS gage 07099500 (and others) Arkansas River Expected Moments Algorithm perception thresholds for the historical and systematic period from 1165 to 2004. This table contains the water year ranges to which each perception threshold applies;  $T_{Y,lower}$ , the lower bound of the perception threshold (in cubic feet per second) for water year  $Y$ ;  $T_{Y,upper}$ , the upper bound of the perception threshold in cfs for water year  $Y$ ; and a comment describing the threshold.

Start year	End year	EMA perception threshold		Comments
		$T_{Y,lower}$	$T_{Y,upper}$	
1165	1858	150,000	Infinity	Paleoflood nonexceedance bound.
1859	1892	40,000	Infinity	1864 Historical information.
1893	1894	19,900	Infinity	1893 Historical information.
1895	1976	0	Infinity	Continuous systematic record.
1977	2004	20,000	Infinity	Post-reservoir bound.

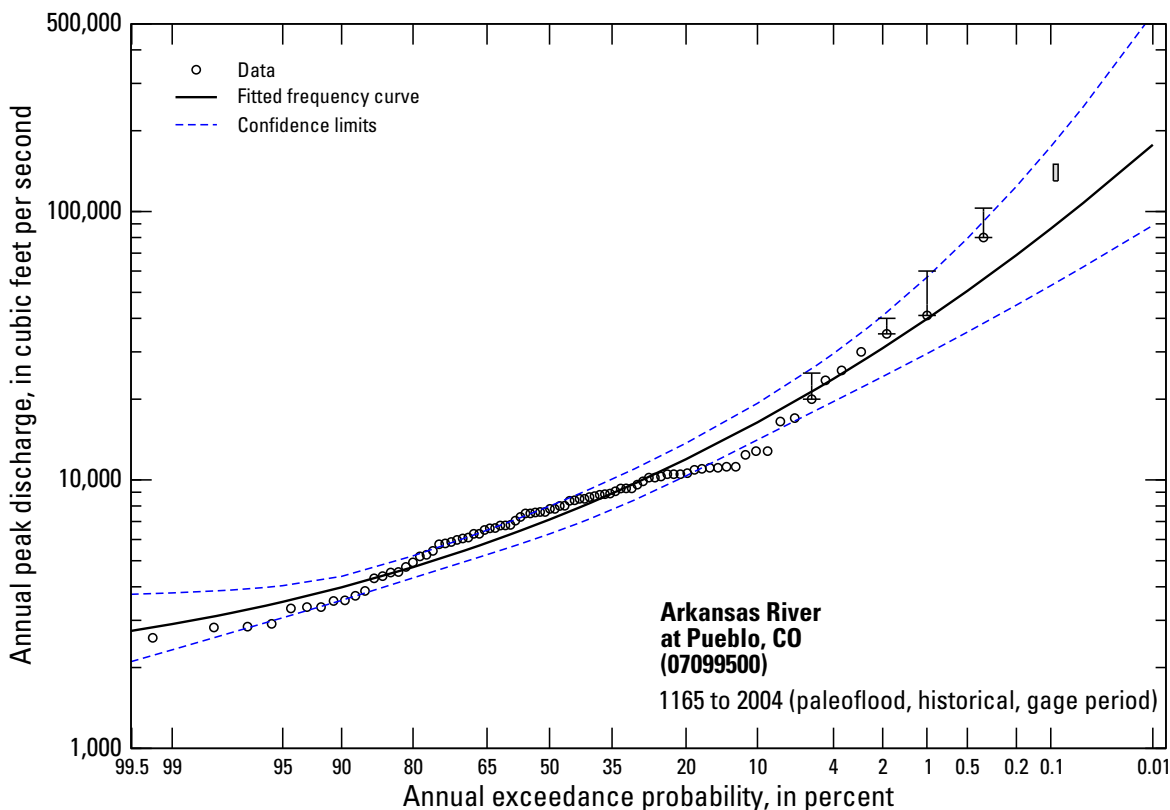
have been estimated (or recorded), had they exceeded approximately 20,000 cfs. Table 10.16 contains the EMA perception thresholds for each water year in the record, including the historical and paleoflood period, for Gage 07099500.

## Results from Flood Frequency Analysis

A flood frequency analysis for the Arkansas River at Pueblo was performed using the EMA flow intervals and perception thresholds as shown in table 10.15 and table 10.16. The output from an at-site flood frequency analysis using EMA with the Multiple Grubbs-Beck test is shown below; no PILFs were identified. Note that station skew was used, thus allowing the focus to be on the at-site data. The fitted frequency curve is displayed in figure 10.9 with estimates provided in table 10.17. The flood frequency results (fig. 10.9) indicate the LP-III model fits the bulk of the data well, including most of the large floods, but underfits the largest flood (June 1921) because of the paleoflood data influence. The paleoflood nonexceedance-bound data at Pueblo State Park increases the peak discharge record length substantially to about 840 years, and has an effect on the upper end of the extrapolated frequency curve principally by reducing the skewness coefficient. One can observe the large positive skew and relatively steep transition between snowmelt-dominant floods to rainfall-dominant floods greater than about 10,000 ft<sup>3</sup>/s. These large rainfall floods are responsible for the shape of the upper portion of the frequency curve. The AEP of the largest flood on record (June 1921) is about 1 in 270 from the exceedance-based plotting position, and about 1 in 1,600 from the LP-III model.

DRAFT: August 31, 2017

---PROVISIONAL---  
 THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
 PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
 PEER REVIEW PLAN.  
 IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
 GEOLOGICAL SURVEY (USGS).  
 IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
 TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.



**Figure 10.9.** Graph showing the peak discharge frequency curve, Arkansas River at Pueblo State Park, including gage, historical, and paleoflood data. Peak discharge estimates from the gage are shown as open circles; vertical bars represent estimated data uncertainty for some of the largest floods. Paleoflood nonexceedance bound shown as a grey box.

**Table 10.17.** Peak-flow quantiles in cubic feet per second for U.S. Geological Survey gage 07099500 (and others) Arkansas River based on flood frequency analysis using Expected Moments Algorithm with Multiple Grubbs-Beck Test; variance of estimate shown in log space.

Annual exceedance probability	EMA estimate	Variance of estimate	Lower 5% confidence limit	Upper 95% confidence limit
0.5	7,100	0.000960	6,300	8,000
0.2	11,900	0.001280	10,400	13,700
0.1	16,400	0.001650	14,100	19,300
0.04	23,800	0.002900	19,600	29,600
0.02	31,000	0.004630	24,300	40,900
0.01	39,800	0.007170	29,500	56,800
0.005	50,600	0.010610	35,600	79,400
0.002	68,800	0.016660	44,800	124,100
0.001	86,300	0.022430	53,000	174,400
0.0001	177,300	0.049590	88,700	545,300

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

## Crest Stage Gage Example—Bear Creek at Ottumwa, Iowa

This example demonstrates how the Expected Moments Algorithm (EMA) and Multiple Grubbs-Beck test (MGBT) can be used to correctly perform a flood frequency analysis when censored data are present with variable perception thresholds from a crest stage gage.

A crest-stage gage (CSG) is a simple, reliable device used to obtain the elevation of the flood peak of a stream. Most commonly, a CSG consists of a vertical metal pipe containing a wood or aluminum staff held in a fixed position with relation to a datum reference. At the bottom of the pipe is a perforated cap containing reggranulated cork. When the water in the stream reaches and exceeds the height of the bottom cap (commonly referred to as the gage base), water is able to enter the pipe. As the water rises up the pipe, the cork floats on the water's surface and as the water reaches its peak and starts to recede, the cork adheres to the staff, thereby retaining the crest stage of the flood (Sauer and Turnipseed, 2010). Thus, CSGs provide a censored record of peak flows, as no annual peak flow that results in a flood stage below the bottom cap of the pipe will be recorded. This example demonstrates how the Expected Moments Algorithm (EMA) with the Multiple Grubbs-Beck test for PILFs can correctly represent these censored annual peak records from CSGs in a flood frequency analysis.

For this example, USGS gage 05489490 Bear Creek at Ottumwa, Iowa is used. This gage is a CSG and has a drainage area of 22.9 square miles. It is located in southeast Iowa in the Southern Iowa Drift Plain land-form region which is characterized by rolling hills and deeply carved stream channels (Prior, 1991). The stream banks and channel bed are comprised of sand, silt, and clay materials that are prone to shifting from hydrologic events. The flood plain areas contain a combination of wooded areas, pasture, and row-crop fields.

USGS gage 05489490 has an annual peak record consisting of 49 peaks, beginning in 1965 and ending in 2014 (Eash and others, 2013, table 1). The annual peaks are listed in table 10.18 (downloaded from USGS NWIS: [http://nwis.waterdata.usgs.gov/nwis/peak/?site\\_no=05489490&agency\\_cd=USGS&](http://nwis.waterdata.usgs.gov/nwis/peak/?site_no=05489490&agency_cd=USGS&)) and shown in figure 10.10.

## EMA Representation of Peak Flow Data for Flood Frequency Analysis

As described in the Data Representation using Flow Intervals and Perception Thresholds section, when using EMA, the annual peak flow for every water year during the historical period is described by a flow interval ( $Q_{Y,lower}$ ,  $Q_{Y,upper}$ ) for each water year  $Y$ . For peaks whose values are known and are not censored, the flow interval can be described as ( $Q_{Y,lower} = Q_Y$ ,  $Q_{Y,upper} = Q_Y$ ). For example, as shown in table 10.18, the peak for the 1965 water year is recorded as 4,000 cfs. This peak is known and is not censored, thus the flow interval for the 1965 water years is ( $Q_{1965,lower} = 4,000$ ,  $Q_{1965,upper} = 4,000$ ).

As shown in table 10.18, there are six censored peaks occurring in 1966, 1971, 1975, 1988, 1997, and 2006. Five of these water years (1966, 1971, 1975, 1997, and 2006) have censored peaks due to the stage of the annual peak not reaching the gage base of the CSG. These peak can be described by flow intervals in which  $Q_{Y,lower} = 0$  and  $Q_{Y,upper} = \text{CSG gage base}$ . Similarly, the annual peak in water year 1988 is censored; however, in this case, the censoring is due to issues related to backwater. The CSG recorded an annual peak of 899 cfs, but it is known that the peak was affected by backwater due to ice causing the recorded peak to be larger than the actual peak. Thus, since there is no further information pertaining to the 1988 peak, it can be represented as a flow interval in which  $Q_{1988,lower} = 0$  and  $Q_{1988,upper} = 899$  cfs. Table 10.19 contains the EMA flow intervals for each water year in the record for gage 05489490.

As described in the Data Representation using Flow Intervals and Perception Thresholds section, EMA distinguishes among sampling properties by employing perception thresholds denoted ( $T_{Y,lower}$ ,  $T_{Y,upper}$ ) for each year  $Y$ , which reflect the range of flows that would have been measured or recorded had they occurred. Perception thresholds describe the range of measurable potential discharges and are independent of the actual

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

**Table 10.18.** U.S. Geological Survey gage 05489490 Bear Creek at Ottumwa, Iowa annual peak-flow record consisting of 49 peaks from 1965 to 2014. This table contains the date of the annual peak recorded at the gage, the water year of the annual peak, and the corresponding annual peak in cubic feet per second (cfs).

Date of peak streamflow	Water year	Annual peak streamflow (cfs)	Date of peak streamflow	Water year	Annual peak streamflow (cfs)	Date of peak streamflow	Water year	Annual peak streamflow (cfs)
1965-09-21	1965	4,000	1982-07-03	1982	4,030	1998-10-05	1999	2,840
1966-00-00	1966	< 1,180	1982-10-08	1983	2,180	2000-06-23	2000	3,520
1967-06-09	1967	2,880	1984-06-08	1984	1,780	2001-05-15	2001	2,430
1967-10-15	1968	1,310	1985-03-04	1985	1,610	2002-05-11	2002	2,670
1968-10-15	1969	1,420	1986-09-19	1986	1,910	2003-06-26	2003	560
1970-06-24	1970	3,130	1987-05-31	1987	990	2004-08-27	2004	3,000
1971-00-00	1971	< 1,180	1988-02-20	1988	< 899	2005-04-12	2005	859
1972-05-08	1972	1,620	1989-09-09	1989	1,820	2006-00-00	2006	< 710
1973-01-19	1973	1,570	1990-05-25	1990	3,120	2007-08-23	2007	2,390
1974-05-19	1974	2,060	1991-04-18	1991	1,850	2008-05-11	2008	3,160
1975-00-00	1975	< 705	1992-09-15	1992	1,840	2009-08-27	2009	2,520
1976-04-24	1976	3,340	1993-05-07	1993	2,410	2010-08-09	2010	3,750
1977-08-07	1977	3,530	1994-06-23	1994	1,400	2011-06-14	2011	2,600
1978-07-21	1978	2,010	1995-04-11	1995	1,560	2012-04-14	2012	1,450
1979-03-29	1979	1,830	1996-05-28	1996	3,130	2013-05-28	2013	3,850
1980-08-17	1980	2,240	1997-00-00	1997	< 714	2014-09-10	2014	1,200
1981-07-04	1981	2,770	1998-06-18	1998	1,940			

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

**Table 10.19.** U.S. Geological Survey gage 05489490 Bear Creek at Ottumwa, Iowa Expected Moments Algorithm flow intervals for the systematic period from 1965 to 2014. This table contains the water year of the annual peak and the corresponding flow interval defined by lower bound,  $Q_{Y,lower}$ , and upper bound,  $Q_{Y,upper}$ , in cubic feet per second (cfs) for each water year  $Y$ .

Water year	$Q_{Y,lower}$	$Q_{Y,upper}$	Comments	Water year	$Q_{Y,lower}$	$Q_{Y,upper}$	Comments
1965	4,000	4,000		1990	3,120	3,120	
1966	0	1,180	Peak < gage base.	1991	1,850	1,850	
1967	2,880	2,880		1992	1,840	1,840	
1968	1,310	1,310		1993	2,410	2,410	
1969	1,420	1,420		1994	1,400	1,400	
1970	3,130	3,130		1995	1,560	1,560	
1971	0	1,180	Peak < gage base.	1996	3,130	3,130	
1972	1,620	1,620		1997	0	714	Peak < gage base.
1973	1,570	1,570		1998	1,940	1,940	
1974	2,060	2,060		1999	2,840	2,840	
1975	0	705	Peak < gage base.	2000	3,520	3,520	
1976	3,340	3,340		2001	2,430	2,430	
1977	3,530	3,530		2002	2,670	2,670	
1978	2,010	2,010		2003	560	560	
1979	1,830	1,830		2004	3,000	3,000	
1980	2,240	2,240		2005	859	859	
1981	2,770	2,770		2006	0	710	Peak < gage base.
1982	4,030	4,030		2007	2,390	2,390	
1983	2,180	2,180		2008	3,160	3,160	
1984	1,780	1,780		2009	2,520	2,520	
1985	1,610	1,610		2010	3,750	3,750	
1986	1,910	1,910		2011	2,600	2,600	
1987	990	990		2012	1,450	1,450	
1988	0	899	Peak affected by backwater.	2013	3,850	3,850	
1989	1,820	1,820		2014	1,200	1,200	

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

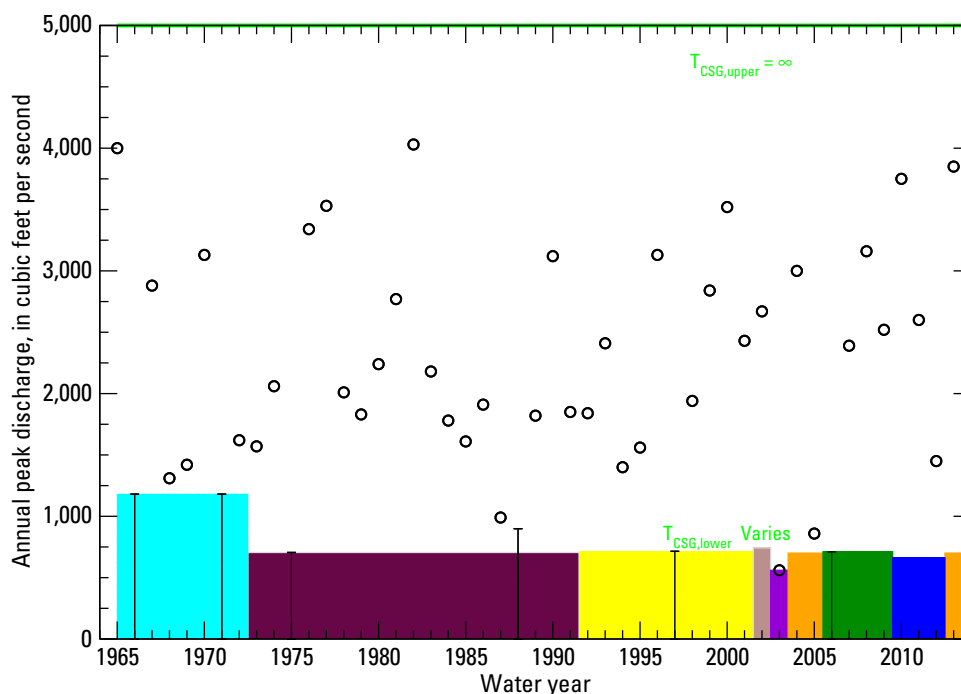
DRAFT: August 31, 2017

peak discharges that have occurred. The lower bound,  $T_{Y,lower}$ , represents the smallest peak flow that would result in a recorded flow in water year  $Y$ . Thus, for a CSG,  $T_{Y,lower}$  can be adjusted to accommodate a changing gage-base discharge. table 10.20 contains the EMA perception thresholds for each water year in the record for Gage 05489490.

The annual peaks, as well as their corresponding EMA flow intervals and perception thresholds, can be displayed graphically. Figure 10.10 shows a representation of the recorded annual peaks, EMA flow intervals, and EMA perception thresholds. The flow intervals whose lower bound is equal to the upper bound are represented by black circles, while the black vertical bars represent the interval flood estimates for those peaks that were not able to be recorded as they were below gage base. The solid-colored blocks represent the many perception thresholds applied to the record. The colored areas represent flows that would be unable to be recorded as they are smaller than the lower bound of the perception threshold  $T_{Y,lower}$ . The white space above the colored areas represents flow ranges for which annual peaks were able to be recorded had they occurred. For example, in figure 10.10, the left-most cyan-colored block represents a perception threshold from 1965 to 1972 where  $T_{Y,lower}=1,180$  cfs,  $T_{Y,upper} = \infty$ . The cyan-colored block spans from 0 cfs to 1,180 cfs, signifying that no annual peak less than 1,180 cfs could be measured during the time period from 1965 to 1972.

## Results from Flood Frequency Analysis

A flood frequency analysis at USGS Gage 05489490 was performed using the EMA flow intervals and perception thresholds as shown in table 10.19 and table 10.20. The output from an at-site flood frequency analysis using EMA with the MGBT to screen for PILFs is shown below. Note that station skew was used, thus allowing the focus to be on the at-site data. The fitted frequency curve is displayed in figure 10.11 with estimates



**Figure 10.10.** Graph showing U.S. Geological Survey gage 05489490 Bear Creek at Ottumwa, Iowa annual peak-flow time series consisting of 49 peaks from 1965 to 2014. The black open circles are the systematic peaks, the black vertical bars with caps represent the interval flood estimates, and the solid rectangle blocks are the perception thresholds.

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

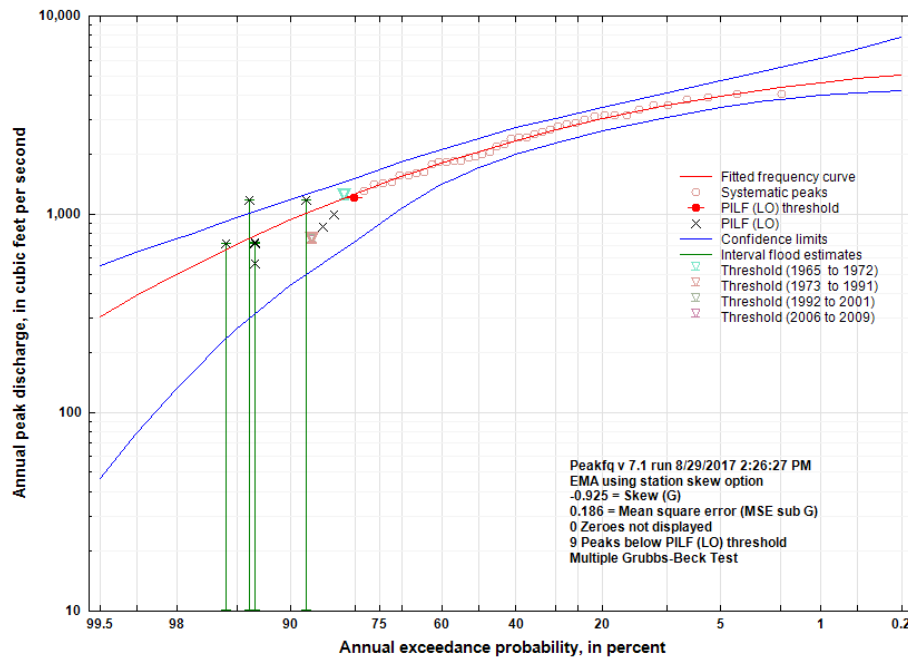


**Table 10.20.** U.S. Geological Survey gage 05489490 Bear Creek at Ottumwa, Iowa Expected Moments Algorithm perception thresholds for the systematic period from 1965 to 2014. This table contains the water year ranges to which each perception threshold applies,  $T_{Y,lower}$  the lower bound of the perception threshold (in cfs) for water year  $Y$ ,  $T_{Y,upper}$ , the upper bound of the perception threshold in cfs for water year  $Y$ , and a comment describing the threshold.

EMA perception threshold				
Start year	End year	$T_{Y,lower}$	$T_{Y,upper}$	Comments
1965	1972	1,180	infinity	Initial gage base of CSG = 1,180 cfs.
1973	1991	705	infinity	Gage base lowered.
1992	2001	714	infinity	Gage base raised as a result of spring thaw.
2002	2002	743	infinity	Gage base raised as a result of spring thaw.
2003	2003	560	infinity	Gage base lowered as a result of routine site visit (HWM).
2004	2005	700	infinity	Gage base raised as a result of spring thaw.
2006	2009	710	infinity	Gage base raised as a result of spring thaw.
2010	2012	661	infinity	Gage base lowered as a result of spring thaw.
2013	2014	700	infinity	Gage base raised as a result of spring thaw.

provided in table 10.21. The final estimated moments were 3.2787 (mean), 0.2331 (standard deviation), and -0.925 (station skew).

As shown in the example above, EMA correctly represents the censored annual peak data through the use



**Figure 10.11.** Graph showing annual exceedance probability plot for USGS Gage 05489490 Bear Creek at Ottumwa, Iowa based on flood frequency analysis using Expected Moments Algorithm with Multiple Grubbs-Beck Test. The red line is the fitted log-Pearson Type III frequency curve, the blue lines are the upper and lower confidence limits, the red circles are the systematic peaks, the green lines represent the interval flood estimates, the solid red circle with a line through it is the potentially influential low floods (PILFs) thresholds as identified by the MGBT, and the black x's are the PILFs identified by the MGBT.

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
 PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
 PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
 GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
 TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

## 160 Guidelines for Determining Flood Flow Frequency – Bulletin 17C

**Table 10.21.** Peak-flow quantiles in cubic feet per second for USGS Gage 05489490 Bear Creek at Ottumwa, Iowa based on flood frequency analysis using Expected Moments Algorithm with Multiple Grubbs-Beck Test; variance of estimate shown in log space.

Annual exceedance probability	EMA estimate	Variance of estimate	Lower 2.5% confidence limit	Upper 97.5% confidence limit
0.5	2,061	0.0012	1,702	2,406
0.2	3,004	0.0009	2,611	3,444
0.1	3,507	0.0008	3,080	4,064
0.04	4,021	0.0010	3,543	4,856
0.02	4,329	0.0013	3,779	5,454
0.01	4,586	0.0018	3,942	6,074
0.005	4,802	0.0024	4,057	6,746
0.002	5,036	0.0033	4,160	7,750

- 1 of flow intervals and perception thresholds. The EMA flow intervals provide a straightforward approach to
- 2 appropriately represent the censored flows, while the perception thresholds accommodate the changing gage
- 3 base. Special thanks to Jon Nania and David Eash of the USGS Iowa Water Science Center for providing data
- 4 and insight relating to USGS Gage 05489490 Bear Creek at Ottumwa, Iowa.

DRAFT: August 31, 2017

---PROVISIONAL---  
 THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
 PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
 PEER REVIEW PLAN.  
 IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
 GEOLOGICAL SURVEY (USGS).  
 IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
 TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

## Historical and PILF Example—Santa Cruz River near Lochiel, Arizona

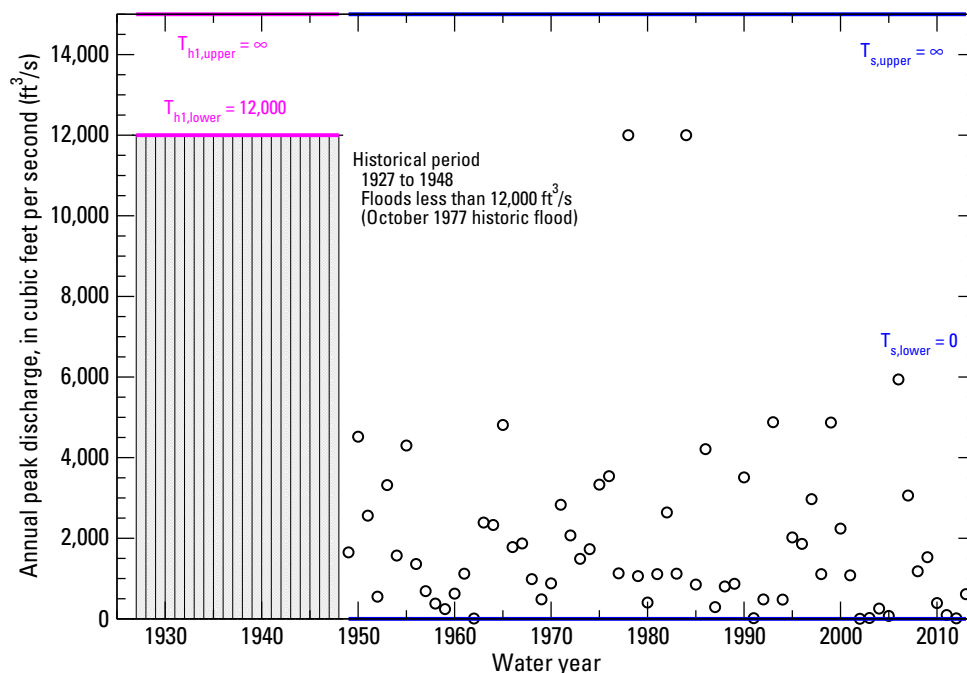
This example illustrates the use of EMA for a historical record with one large flood (described in the section Historical Flood Information) and a number of PILFs (described in section Zero Flows and **Potentially-Influential Potentially Influential** Low Floods).

For this example, USGS gage 09480000 Santa Cruz River near Lochiel, Arizona, is used. The Santa Cruz River is a tributary to the Gila River; the 82.2-square-mile watershed lies within the Basin and Range province in Arizona (Paretti and others, 2014a). This gaging station was previously analyzed by Cohn and others (2017) and Paretti and others (2014a, fig. 21).

Gage 09480000 has an annual peak record consisting of 65 peaks, beginning in 1949 and ending in 2013. There is a historic flood that occurred within the period of gaging record on October 9, 1977. This flood is noted in the USGS Annual Water Data Report for this gage, available in the peak-flow file, and there is historical information available for this large flood (Aldridge and Eychaner, 1984) that indicates this flood is the largest since 1927. The annual peaks are listed in table 10.22 and shown in figure 10.12. Of the 65 annual peaks, the August 15, 1984 flood is equal to the October 1977 historic flood peak. Based on the historical flood information in Aldridge and Eychaner (1984) for the 1977 flood, information from the October 1977 flood is used as a perception threshold to represent the 22 years of missing information from 1927–1948.

## EMA Representation of Peak Flow Data for Flood Frequency Analysis

As described in the Data Representation using Flow Intervals and Perception Thresholds section, when using EMA, the annual peak flow for every water year during the historical period is described by a flow interval ( $Q_{Y,lower}$ ,  $Q_{Y,upper}$ ) for each water year  $Y$ . For peaks whose values are known and are not censored, the flow



**Figure 10.12.** Graph showing U.S. Geological Survey gage 09480000 Santa Cruz River near Lochiel, Arizona annual peak-flow time series consisting of 65 peaks from 1949 to 2013. The historical period is shown as a grey shaded area, with a perception threshold (12,000 cfs) estimated from the October 1977 historic flood.

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

**Table 10.22.** U.S. Geological Survey gage 09480000 Santa Cruz River near Lochiel, Arizona annual peak-flow record consisting of 65 peaks from 1949 to 2013. This table contains the date of the annual peak recorded at the gage, the water year of the annual peak, and the corresponding annual peak in cubic feet per second (cfs).

Date of peak streamflow	Water year	Annual peak streamflow (cfs)	Date of peak streamflow	Water year	Annual peak streamflow (cfs)	Date of peak streamflow	Water year	Annual peak streamflow (cfs)
1949-09-13	1949	1,650	1971-08-10	1971	2,830	1993-01-18	1993	4,880
1950-07-30	1950	4,520	1972-07-16	1972	2,070	1994-08-30	1994	478
1951-08-02	1951	2,560	1973-06-30	1973	1,490	1995-07-12	1995	2,020
1952-08-16	1952	550	1974-08-04	1974	1,730	1996-07-10	1996	1,860
1953-07-14	1953	3,320	1975-07-22	1975	3,330	1997-09-11	1997	2,970
1954-07-22	1954	1,570	1976-07-22	1976	3,540	1998-07-07	1998	1,110
1955-08-06	1955	4,300	1977-09-05	1977	1,130	1999-07-28	1999	4,870
1956-07-17	1956	1,360	1977-10-09	1978	12,000	2000-08-06	2000	2,240
1957-08-09	1957	688	1979-01-25	1979	1,060	2000-10-22	2001	1,080
1958-08-07	1958	380	1980-06-30	1980	406	2002-03-04	2002	1.5
1959-08-14	1959	243	1981-07-15	1981	1,110	2003-08-14	2003	22
1960-07-30	1960	625	1982-08-11	1982	2,640	2004-08-05	2004	256
1961-08-08	1961	1,120	1983-03-04	1983	1,120	2005-08-23	2005	73
1962-07-29	1962	7.6	1984-08-15	1984	12,000	2006-08-08	2006	5,940
1963-08-25	1963	2,390	1985-07-19	1985	850	2007-07-19	2007	3,060
1964-09-09	1964	2,330	1986-08-29	1986	4,210	2008-07-23	2008	1,180
1965-09-12	1965	4,810	1987-08-10	1987	291	2009-07-20	2009	1,530
1966-08-18	1966	1,780	1988-08-23	1988	804	2010-07-31	2010	392
1967-08-03	1967	1,870	1989-08-04	1989	871	2011-08-13	2011	95
1967-12-20	1968	986	1990-07-17	1990	3,510	2012-07-28	2012	12
1969-08-05	1969	484	1991-07-26	1991	17	2013-09-08	2013	612
1970-08-03	1970	880	1992-08-01	1992	483			

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

interval can be described as ( $Q_{Y,lower} = Q_Y$ ,  $Q_{Y,upper} = Q_Y$ ). In this example, the flow values are known for all the years where the gage was in operation. Table 10.23 contains the EMA flow intervals for each water year in the record for gage 09480000. The historical period is described by a perception threshold.

As described in the Data Representation using Flow Intervals and Perception Thresholds section, EMA distinguishes among sampling properties by employing perception thresholds denoted ( $T_{Y,lower}$ ,  $T_{Y,upper}$ ) for each year  $Y$ , which reflect the range of flows that would have been measured or recorded had they occurred. Perception thresholds describe the range of measurable potential discharges and are independent of the actual peak discharges that have occurred. The lower bound,  $T_{Y,lower}$ , represents the smallest peak flow that would result in a recorded flow in water year  $Y$ . For most peaks at most gages,  $T_{Y,upper}$  is assumed to be infinite, as bigger floods that might exceed the measurement capability of the streamgage are determined through study of high-water marks and other physical evidence of the flood. For periods of continuous, full-range peak-flow record, the perception threshold is represented by ( $T_{Y,lower} = 0$ ,  $T_{Y,upper} = \infty$ ), where  $T_{Y,lower} = 0$  is the gage-base discharge. Based on the October 1977 large historical flood (Aldridge and Eychaner, 1984), it is known that floods at this location would have been estimated (or recorded) had they exceeded approximately 12,000 cfs. Table 10.24 contains the EMA perception thresholds for each water year in the record, including the historical period, for gage 09480000.

## Results from Flood Frequency Analysis

A flood frequency analysis at USGS gage 09480000 was performed using the EMA flow intervals and perception thresholds as shown in table 10.23 and table 10.24. The output from an at-site flood frequency analysis using EMA with the Multiple Grubbs-Beck test to screen for PILFs is shown below. Note that station skew was used, thus allowing the focus to be on the at-site data. The fitted frequency curve is displayed in figure 10.13 with estimates provided in table 10.25. The final estimated moments were 3.0691 (mean), 0.4898 (standard deviation), and -0.462 (station skew).

As shown in figure 10.13, there are two floods that exceed the historical threshold (12,000 cfs): the October 1977 flood and the August 1984 flood. Using MGBT, ten PILFs were identified, with a threshold equal to 380 ft<sup>3</sup>/s and a significance level equal to 0.0228. Thus, all 10 annual peaks less than 380 ft<sup>3</sup>/s are censored and recoded in the framework of EMA with flow intervals of ( $Q_{Y,lower} = 0$ ,  $Q_{Y,upper} = 380$ ). The MGBT threshold also has the effect of adjusting the lower bound of the perception threshold. Thus, for the systematic period from 1949 to 2013, the perception threshold is ( $T_{Y,lower} = 380$ ,  $T_{Y,upper} = \infty$ ). For the historical information, the lower threshold is ( $T_{Y,lower} = 12,000$ ) (table 10.24). As shown in figure 10.13, by censoring the eight smallest peaks in the record, the remaining smallest annual exceedance probability peaks and the largest floods are well fit by the frequency curve (red line).

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

**Table 10.23.** U.S. Geological Survey gage 09480000 Santa Cruz River near Lochiel, Arizona Expected Moments Algorithm flow intervals for the systematic period from 1949 to 2013. This table contains the water year of the annual peak and the corresponding flow interval defined by lower bound,  $Q_{Y,lower}$ , and upper bound,  $Q_{Y,upper}$ , in cubic feet per second (cfs) for each water year  $Y$ .

Water year	$Q_{Y,lower}$	$Q_{Y,upper}$	Comments	Water year	$Q_{Y,lower}$	$Q_{Y,upper}$	Comments
1949	1,650	1,650		1982	2,640	2,640	
1950	4,520	4,520		1983	1,120	1,120	
1951	2,560	2,560		1984	12,000	12,000	
1952	550	550		1985	850	850	
1953	3,320	3,320		1986	4,210	4,210	
1954	1,570	1,570		1987	291	291	
1955	4,300	4,300		1988	804	804	
1956	1,360	1,360		1989	871	871	
1957	688	688		1990	3,510	3,510	
1958	380	380		1991	17	17	
1959	243	243		1992	483	483	
1960	625	625		1993	4,880	4,880	
1961	1,120	1,120		1994	478	478	
1962	7.6	7.6		1995	2,020	2,020	
1963	2,390	2,390		1996	1,860	1,860	
1964	2,330	2,330		1997	2,970	2,970	
1965	4,810	4,810		1998	1,110	1,110	
1966	1,780	1,780		1999	4,870	4,870	
1967	1,870	1,870		2000	2,240	2,240	
1968	986	986		2001	1,080	1,080	
1969	484	484		2002	1.5	1.5	
1970	880	880		2003	22	22	
1971	2,830	2,830		2004	256	256	
1972	2,070	2,070		2005	73	73	
1973	1,490	1,490		2006	5,940	5,940	
1974	1,730	1,730		2007	3,060	3,060	
1975	3,330	3,330		2008	1,180	1,180	
1976	3,540	3,540		2009	1,530	1,530	
1977	1,130	1,130		2010	392	392	
1978	12,000	12,000	Historic flood.	2011	95	95	
1979	1,060	1,060		2012	12	12	
1980	406	406		2013	612	612	
1981	1,110	1,110					

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

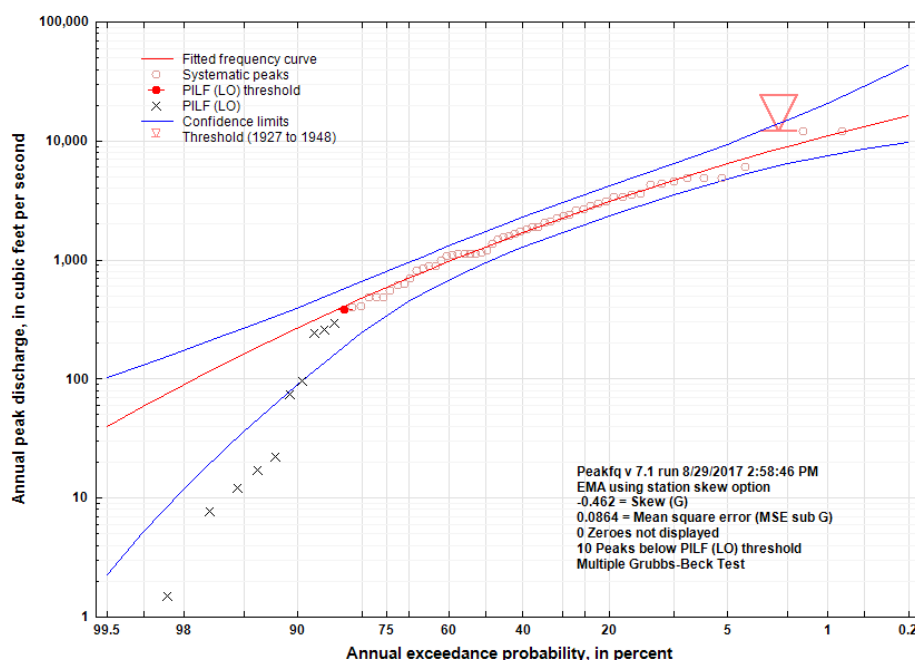
IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

**Table 10.24.** U.S. Geological Survey gage 09480000 Santa Cruz River near Lochiel, Arizona Expected Moments Algorithm perception thresholds for the historical and systematic period from 1927 to 2013. This table contains the water year ranges to which each perception threshold applies,  $T_{Y,lower}$  the lower bound of the perception threshold (in cfs) for water year  $Y$ ,  $T_{Y,upper}$ , the upper bound of the perception threshold in cfs for water year  $Y$ , and a comment describing the threshold.

Start year	End Year	EMA perception threshold		Comments
		$T_{Y,lower}$	$T_{Y,upper}$	
1927	1948	12,000	infinity	Historical information.
1949	2013	0	infinity	Continuous systematic record.



**Figure 10.13.** Graph showing annual exceedance probability plot for USGS Gage 09480000 Santa Cruz River near Lochiel, Arizona based on flood frequency analysis using Expected Moments Algorithm with Multiple Grubbs-Beck Test. The red line is the fitted log-Pearson Type III frequency curve, the blue lines are the upper and lower confidence limits, the black circles are the systematic peaks, the solid red circle with a line through it is the potentially influential low floods (PILFs) thresholds as identified by the MGBT, and the black x's are the PILFs identified by the MGBT.

DRAFT: August 31, 2017

---PROVISIONAL---  
THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.  
IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).  
IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.



**Table 10.25.** Peak-flow quantiles in cubic feet per second for USGS Gage 09480000 Santa Cruz River near Lochiel, Arizona based on flood frequency analysis using Expected Moments Algorithm with Multiple Grubbs-Beck Test; variance of estimate shown in log space.

Annual exceedance probability	EMA estimate	Variance of estimate	Lower 2.5% confidence limit	Upper 97.5% confidence limit
0.5	1,279	0.0042	936.1	1,719
0.2	3,079	0.0040	2,314	4,138
0.1	4,652	0.0042	3,481	6,394
0.04	6,982	0.0056	5,119	10,460
0.02	8,914	0.0076	6,337	14,780
0.01	10,970	0.0105	7,474	20,570
0.005	13,150	0.0144	8,509	28,280
0.002	16,170	0.0211	9,719	42,560

DRAFT: August 31, 2017

---PROVISIONAL---  
 THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
 PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
 PEER REVIEW PLAN.  
 IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
 GEOLOGICAL SURVEY (USGS).  
 IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
 TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

## Paleoflood Record Example—American River at Fair Oaks, California

This example illustrates the use of EMA for a historical record with several large floods (described in the Section Historical Flood Information) and detailed paleoflood data (described in the Section Paleoflood and Botanical Information), utilizing multiple censoring and interval data for Reclamation's Folsom Dam (Bureau of Reclamation, 2002). For this example, USGS gage 11446500 American River at Fair Oaks, California, is used, with additional data from the historical and paleoflood record, interpretations from USACE and Bureau of Reclamation data and investigations, and other historical sources. In 1986 and 1997, floods on the American River and in central California heightened concerns about the hydrologic risk at Folsom Dam. In part, these concerns led to two National Research Council panels to evaluate American River flood hazards (National Research Council, 1995, 1999). These panels reviewed flood control and flood-plain-management issues, focusing on estimating floods with AEPs greater than 0.005 (1 in 200), specifically 1 in 100 (0.01). For dam safety, the primary concern is floods with very small AEPs generally in the range of 0.001 to 0.0001 (1 in 1,000 to 1 in 10,000). For this example, these estimates are made using gage, historical, and paleoflood data. This example illustrates the use of a very long paleoflood record with large floods outside the gaging period, historical information, multiple thresholds, and interval observations. This example is meant to be illustrative and is not intended to be used for making flood-plain-management decisions along the American River.

Bureau of Reclamation (2002) conducted a paleoflood and flood frequency study to investigate these issues. The primary objective of the study was to develop an estimate of peak discharge frequency of the American River at Folsom Dam in the above annual probability range. The peak discharge frequency information was subsequently combined with historical hydrographs to develop probabilistic hydrographs based on paleoflood information. Paleoflood information for the Bureau of Reclamation (2002) study was based on geomorphic, stratigraphic, and geochronologic information collected from four sites in the American River basin: (1) South Fork American River near Kyburz, (2) South Fork American River near Lotus, (3) North Fork of the American River at Ponderosa Bridge, and (4) lower American River near Fair Oaks. Two main types of paleoflood data were collected from the four sites to evaluate the flood hazard for Folsom Dam: (1) paleoflood magnitude and age estimates for the South Fork near Kyburz and Lotus, and the lower American River, and (2) a single paleohydrologic bound for the North Fork. Stratigraphic information from 14 sites provides evidence for late Holocene paleofloods that are preserved at or above the peak stage of the largest historical floods. The age of these paleofloods is constrained by 38 radiocarbon ages, published archaeological age correlations, and published obsidian hydration age estimates.

For this example, peak-flow data from the lower American River at Fair Oaks, California (Gage 11446500) are used and modified based on additional information from USGS Water-Supply Papers, USACE records, Bureau of Reclamation records and investigations, and other historical information. As such, they do not directly correspond to peak flows in the USGS NWIS database. There are 77 peaks beginning in 1905 and ending in 1997, with several years with very low floods or missing values (1910, 1912–13, 1918, 1929, 1977). Large historical floods occurred in 1997, 1986, and 1862, and are described in National Research Council (1999) and Bureau of Reclamation (2002). The paleoflood period covers the past 2,000 years, from year 1 to 1847, the historical period begins in 1848, and the gaging period begins in 1905. The annual peaks, historical floods, and paleofloods are listed in Table 10.26 and shown in Figure 10.14. Perception thresholds are estimated based on the March 1907 flood, the January 1862 flood, and paleofloods.

## EMA Representation of Peak Flow Data for Flood Frequency Analysis

As described in the Data Representation using Flow Intervals and Perception Thresholds Section, when using EMA, the annual peak flow for every water year during the historical period is described by a flow interval

**DRAFT: August 31, 2017**

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.  
IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).  
IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

**Table 10.26.** U.S. Geological Survey gage 11446500 American River at Fair Oaks annual peak-flow record consisting of 77 peaks from 1905 to 1997, with historical floods and paleofloods. Horizontal lines indicate breaks in data.

Water year	Annual peak streamflow (cfs)	Water year	Annual peak streamflow (cfs)	Water year	Annual peak streamflow (cfs)
650	>600,000	1933	16,500	1961	8,000
1437	>400,000	1934	22,600	1962	40,000
1574	>400,000	1935	60,900	1963	240,000
1711	>400,000	1936	58,300	1964	24,000
1862	>262,000	1937	33,000	1965	260,000
1905	24,200	1938	114,000	1966	6,500
1906	59,700	1939	10,900	1967	46,000
1907	156,000	1940	89,200	1968	30,000
1908	10,300	1941	38,800	1969	120,000
1909	119,000	1942	83,200	1970	122,000
1911	81,300	1943	152,000	1971	48,000
1914	74,100	1944	20,100	1972	12,000
1915	47,900	1945	94,400	1973	69,000
1916	40,700	1946	42,200	1974	55,000
1917	42,300	1947	27,900	1975	46,000
1919	67,500	1948	21,000	1976	15,000
1920	20,100	1949	37,500	1978	40,000
1921	39,200	1950	34,400	1979	33,000
1922	31,600	1951	180,000	1980	175,000
1923	39,000	1952	37,200	1981	20,000
1924	14,000	1953	49,700	1982	152,000
1925	99,500	1954	42,600	1983	93,000
1926	27,400	1955	10,800	1984	88,000
1927	67,700	1956	219,000	1985	17,000
1928	163,000	1957	42,000	1986	259,000
1930	24,400	1958	54,000	1997	298,000
1931	9,900	1959	20,000		
1932	21,100	1960	75,000		

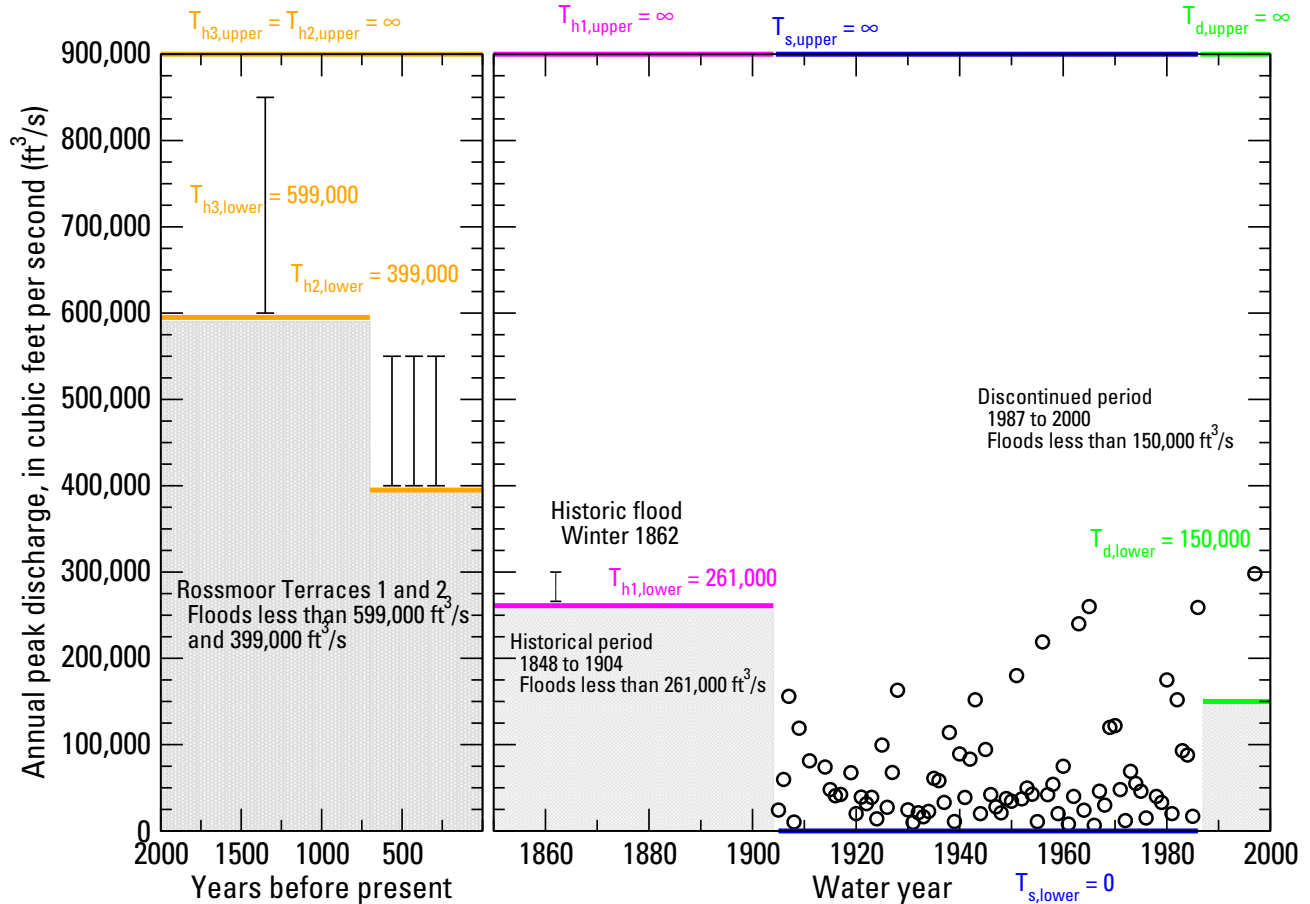
---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017



**Figure 10.14.** Graph showing approximate unregulated peak discharge and paleoflood estimates, with historical and paleoflood exceedance thresholds, American River at Fair Oaks. A scale break is used to separate the gaging station data from the much longer paleoflood record. Mean values of paleofloods threshold age data are plotted for simplicity.

$(Q_{Y,lower}, Q_{Y,upper})$  for each water year  $Y$ . For peaks whose values are known and are not censored, the flow interval can be described as  $(Q_{Y,lower} = Q_Y, Q_{Y,upper} = Q_Y)$ . In this example, the flow values are known for all the years where the gage was in operation. Table 10.27 contains the EMA flow intervals for each water year in the record for gage 11446500. The historical and paleoflood periods are described by perception thresholds.

As described in the Data Representation using Flow Intervals and Perception Thresholds Section, EMA distinguishes among sampling properties by employing perception thresholds denoted  $(T_{Y,lower}, T_{Y,upper})$  for each year  $Y$ , which reflect the range of flows that would have been measured or recorded had they occurred. Perception thresholds describe the range of measurable potential discharges and are independent of the actual peak discharges that have occurred. The lower bound,  $T_{Y,lower}$ , represents the smallest peak flow that would result in a recorded flow in water year  $Y$ . For most peaks at most gages,  $T_{Y,upper}$  is assumed to be infinite, as bigger floods that might exceed the measurement capability of the streamgage are determined through study of highwater marks and other physical evidence of the flood. For periods of continuous, full-range peak flow record, the perception threshold is represented by  $(T_{Y,lower} = 0, T_{Y,upper} = \infty)$ , where  $T_{Y,lower} = 0$  is the gage-base discharge. Based on the March 1907 large historical flood (Bureau of Reclamation, 2002), it is known that floods at this location would have been estimated (or recorded), had they exceeded approximately 150,000 cfs. Table 10.28 contains the EMA perception thresholds for each water year in the record, including the historical period, for gage 11446500.

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

## 170 Guidelines for Determining Flood Flow Frequency – Bulletin 17C

**Table 10.27.** U.S. Geological Survey gage 11446500 American River at Fair Oaks Expected Moments Algorithm flow intervals for the historical and systematic period from 650 to 1997.

Water year	Q <sub>Y,lower</sub>	Q <sub>Y,upper</sub>	Comments	Water year	Q <sub>Y,lower</sub>	Q <sub>Y,upper</sub>	Comments
650	600,000	850,000	Paleoflood.	1946	42,200	42,200	
1437	400,000	550,000	Paleoflood.	1947	27,900	27,900	
1574	400,000	550,000	Paleoflood.	1948	21,000	21,000	
1711	400,000	550,000	Paleoflood.	1949	37,500	37,500	
1862	262,000	300,000	Historical flood.	1950	34,400	34,400	
1905	24,200	24,200		1951	180,000	180,000	
1906	59,700	59,700		1952	37,200	37,200	
1907	156,000	156,000		1953	49,700	49,700	
1908	10,300	10,300		1954	42,600	42,600	
1909	119,000	119,000		1955	10,800	10,800	
1911	81,300	81,300		1956	219,000	219,000	
1914	74,100	74,100		1957	42,000	42,000	
1915	47,900	47,900		1958	54,000	54,000	
1916	40,700	40,700		1959	20,000	20,000	
1917	42,300	42,300		1960	75,000	75,000	
1919	67,500	67,500		1961	8,000	8,000	
1920	20,100	20,100		1962	40,000	40,000	
1921	39,200	39,200		1963	240,000	240,000	
1922	31,600	31,600		1964	24,000	24,000	
1923	39,000	39,000		1965	260,000	260,000	
1924	14,000	14,000		1966	6,500	6,500	
1925	99,500	99,500		1967	46,000	46,000	
1926	27,400	27,400		1968	30,000	30,000	
1927	67,700	67,700		1969	120,000	120,000	
1928	163,000	163,000		1970	122,000	122,000	
1930	24,400	24,400		1971	48,000	48,000	
1931	9,900	9,900		1972	12,000	12,000	
1932	21,100	21,100		1973	69,000	69,000	
1933	16,500	16,500		1974	55,000	55,000	
1934	22,600	22,600		1975	46,000	46,000	
1935	60,900	60,900		1976	15,000	15,000	
1936	58,300	58,300		1978	40,000	40,000	
1937	33,000	33,000		1979	33,000	33,000	
1938	114,000	114,000		1980	175,000	175,000	
1939	10,900	10,900		1981	20,000	20,000	
1940	89,200	89,200		1982	152,000	152,000	
1941	38,800	38,800		1983	93,000	93,000	
1942	83,200	83,200		1984	88,000	88,000	
1943	152,000	152,000		1985	17,000	17,000	---PROVISIONAL---
1944	20,100	20,100		1986	259,000	259,000	THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS PEER REVIEW PLAN.
1945	94,400	94,400		1997	298,000	298,000	IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S. GEOLOGICAL SURVEY (USGS).

DRAFT: August 31, 2017

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

**Table 10.28.** U.S. Geological Survey gage 11446500 American River at Fair Oaks Expected Moments Algorithm perception thresholds for the paleoflood, historical, and systematic period from 1 to 2000. This table contains the water year ranges to which each perception threshold applies,  $T_{Y,lower}$  the lower bound of the perception threshold (in cfs) for water year  $Y$ ,  $T_{Y,upper}$  the upper bound of the perception threshold in cfs for water year  $Y$ , and a comment describing the threshold.

Start year	End year	EMA perception threshold		Comments
		$T_{Y,lower}$	$T_{Y,upper}$	
1	1301	599,000	infinity	Lower Rossmoor terrace 1.
1302	1847	399,000	infinity	Lower Rossmoor terrace 2.
1848	1904	261,000	infinity	1862 historical threshold.
1905	1909	0	infinity	Gage record.
1910	1910	150,000	infinity	March 1907 low floods and missing.
1911	1911	0	infinity	Gage record.
1912	1913	150,000	infinity	March 1907 low floods and missing.
1914	1917	0	infinity	Gage record.
1918	1918	150,000	infinity	March 1907 low floods and missing.
1919	1928	0	infinity	Gage record.
1929	1929	150,000	infinity	March 1907 low floods and missing.
1930	1976	0	infinity	Gage record.
1977	1977	150,000	infinity	March 1907 low floods and missing.
1978	1986	0	infinity	Gage record.
1987	1996	150,000	infinity	March 1907 low floods and missing.
1997	1997	0	infinity	Gage record.
1998	2000	150,000	infinity	March 1907 low floods and missing.

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

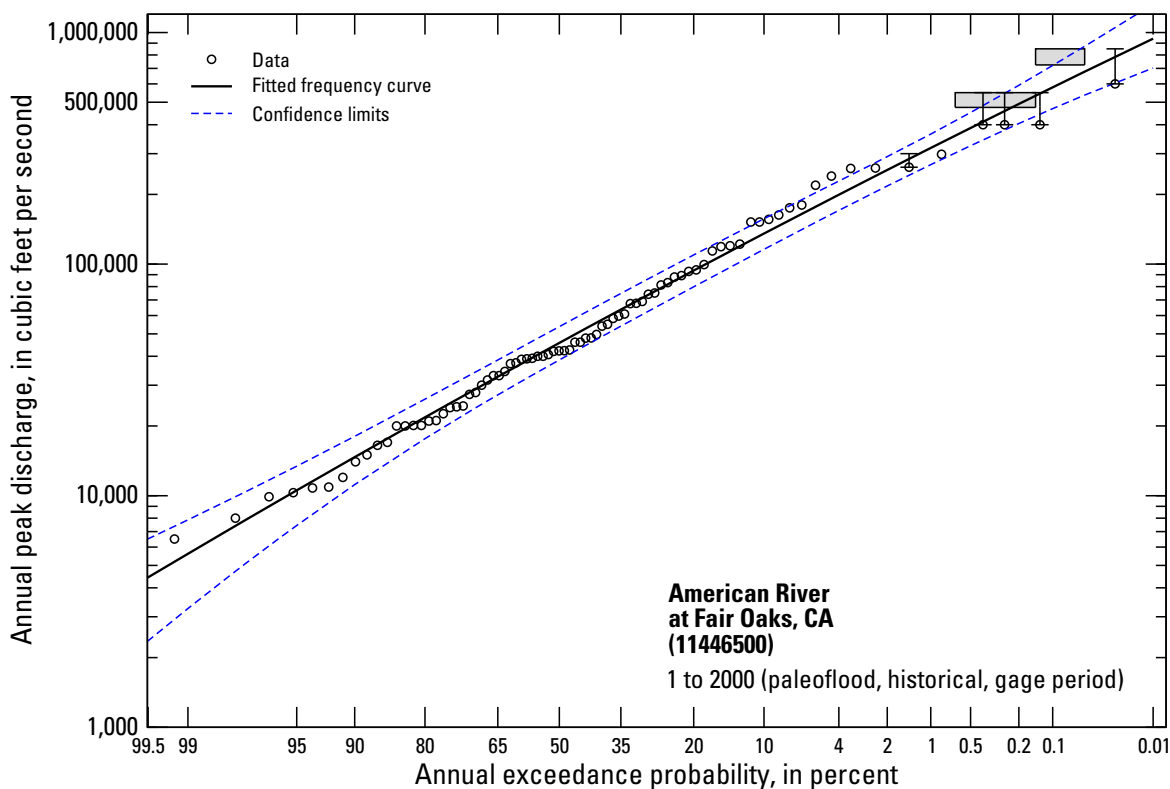
IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

## Results from Flood Frequency Analysis

A flood frequency analysis at USGS Gage 11446500 was performed using the EMA flow intervals and perception thresholds as shown in Table 10.27 and Table 10.28. The output from an at-site flood frequency analysis using EMA with the Multiple Grubbs-Beck test to screen for PILFs is shown below. Note that station skew was used, thus allowing the focus to be on the at-site data. The fitted frequency curve is displayed in Figure 10.15 with estimates provided in Table 10.29. Peak discharge estimates for the interval floods are shown in the figure with estimated uncertainty. Peak discharge probabilities are estimated using Cunnane's plotting position with the threshold-exceedance formula that includes paleoflood data. The results indicate that the LP-III model provides an adequate fit to the gage and paleoflood data.



**Figure 10.15.** Graph showing approximate unregulated peak discharge frequency curve, American River at Fair Oaks, including gage, historical and paleoflood data. Peak discharge estimates from the gage are shown as open circles; vertical bars represent estimated data uncertainty for some of the largest floods. Paleoflood nonexceedance bounds shown as grey boxes.

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017



**Table 10.29.** Peak-flow quantiles in cubic feet per second for USGS Gage 11446500 American River at Fair Oaks based on flood frequency analysis using Expected Moments Algorithm with Multiple Grubbs-Beck Test; variance of estimate shown in log space.

Annual exceedance probability	EMA estimate	Variance of estimate	Lower 2.5% confidence limit	Upper 97.5% confidence limit
0.5	45,700	0.001890	38,600	53,700
0.2	93,800	0.001730	79,800	109,600
0.1	135,500	0.001590	115,800	157,000
0.04	199,400	0.001460	170,700	228,300
0.02	255,000	0.001450	217,500	291,100
0.01	317,500	0.001570	268,800	364,500
0.005	387,300	0.001830	324,600	451,400
0.002	491,600	0.002440	404,700	591,900
0.001	580,200	0.003110	469,300	720,800
0.0001	941,200	0.006810	702,800	1,325,300

DRAFT: August 31, 2017

---PROVISIONAL---  
 THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
 PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
 PEER REVIEW PLAN.  
 IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
 GEOLOGICAL SURVEY (USGS).  
 IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
 TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

- Advisory Committee on Water Information.–  
 Annual Exceedance Probability.–  
 Annual Maximum Series.–  
 Bayesian Generalized Least Squares.–  
 Cumulative Distribution Function.–  
 Crest-stage gage.–  
 Expected Moments Algorithm.–  
 El Niño–Southern Oscillation.–  
 Federal Emergency Management Agency.–  
 Generalized Least Squares.–  
 Hydrologic Frequency Analysis Work Group.–  
 High-water mark.–  
 Log-Pearson Type III distribution.–  
 Multiple Grubbs-Beck Test.–  
 Maintenance of Variance Extension.–  
 Mean-Square Error.–  
 Natural Resources Conservation Service.–  
 USGS National Water Information System.–  
 National Weather Service.–  
 Ordinary Least Squares.–  
 Partial-Duration Series.–  
 Potentially-Influential Low Flood.–  
 Paleostage indicator.–  
 Bureau of Reclamation.–  
 River Forecast Center (NWS).–  
 Subcommittee on Hydrology.–  
 U.S. Army Corps of Engineers.–  
 Weighted Least Squares.–  
 Plotting position parameter that is dependent on an assumed distribution ( $0 \leq a \leq 0.5$ );  $a = 0 \equiv$  Weibull;  $a = 0.5 \equiv$  Hazen (eq. 2).–  
 Number of floods/records that exceed a censoring level, such as a historical threshold  $T_h$  or low-outlier threshold, during the systematic record  $n_s$  ( $e_s \leq k$ ;  $e_s < n_s$ ).–  
 Number of floods/records that exceed the historical threshold  $T_h$  during the historical/paleoflood period  $n_h$  ( $e_h \leq k$ ;  $e_h < n_h$ ).–  
 Total number of known flood (observations) during the entire period of observation record  $n$  ( $g = n_s + k - e_s = n_s + e_h$ ).–  
 At-site (station) sample skew coefficient (in log space).–  
 Regional sample skew coefficient (in log space).–  
 Weighted skew coefficient.–  
 Total number of floods/records that exceed a censoring level, such as a historical threshold  $T_h$  or low-outlier threshold, during the entire period of observation record  $n$  ( $k = e_s + e_h$ ).–  
 At-site (station) sample mean (in log space).–  
 At-site (station) sample standard deviation (in log space).–  
 Total peak-flow period of record (years), including systematic  $n_s$  and historical  $n_h$  periods, as available, where  $n = n_s + n_h$ .–  
 Length of the historical period (years); possibly includes a paleoflood period ( $n_h < n$ ).–

DRAFT: August 31, 2017

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
 PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
 PEER REVIEW PLAN.  
 IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
 GEOLOGICAL SURVEY (USGS).  
 IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
 TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

- Length of the peak-flow systematic (gaging) record (years) ( $n_s \leq n$ ). 1
- Annual exceedance probability ( $p$ ),  $p = 1 - q$ . 2
- Cumulative probability,  $q = 1 - p$ . 3
- Flood discharge. 4
- Base discharge. Can be a constant, or vary with each year at a gaging station or CSG. 5
- Discharge quantile for annual exceedance probability  $p$ . 6
- Discharge quantile for cumulative probability  $q$ , equivalent to  $Q_p$ . 7
- Flood discharge estimate in year  $Y$ . 8
- Discharge lower bound for year  $Y$  in EMA. 9
- Discharge upper bound for year  $Y$  in EMA. 10
- Perception threshold for a discontinued (broken) record at a gaging station with systematic period  $n_s$ . 11
- Perception threshold for a historical period  $n_h$ . 12
- PILF censoring threshold from the. 13
- Perception threshold lower bound for Year  $Y$  in EMA; represents the smallest peak flow that would result in a recorded flow. 14
- Perception threshold upper bound for Year  $Y$  in EMA; represents the largest peak flow that would result in a recorded flow. 15
- Base-10 logarithm of a perception threshold for a historical period  $n_h$ . 16
- Base-10 logarithm of the PILF censoring threshold from the. 17
- Base-10 logarithm of discharge lower bound  $Q_{Y,lower}$  for Year  $Y$  in EMA. 18
- Base-10 logarithm of discharge upper bound  $Q_{Y,upper}$  for Year  $Y$  in EMA. 19
- Year. 20
- The probability that flooding will occur in any given year considering the full range of possible annual floods. 21
- The highest instantaneous peak discharge in each year of record. Practically, this is the highest value observed in the record of 15 minute or 60 minute values, depending on the recording interval of the device. Sometimes the maximum mean daily discharge is used on larger rivers. 22
- A list of annual maximum floods. 23
- A general term for a set of any kind of data in which each item is the maximum or minimum in a year. 24
- The presence of autocorrelation indicates that the data in the time series are not random. Rather, future values are correlated with past values. Autocorrelation is calculated as the correlation between the values in a time series and the values in that same time series lagged by one or more timesteps (i.e., the correlation between  $X_i$  and  $X_{i+k}$  where  $i$  is the timestep and  $k$  is the lag). Also known as serial correlation. 25
- A discharge value, determined for selected stations, above which peak discharge data are published. The base discharge at each station is selected so that an average of about three peak flows per year will be published (Langbein and Iseri, 1960). 26
- Floods that exceeded a threshold, where one knows only that a flood was larger than some level, and does not know the magnitude of the flood (Russell, 1982; Stedinger and Cohn, 1986). 27
- A systematic record which is divided into separate continuous segments because of deliberate discontinuation of recording for significant periods of time. This typically occurs when a gage is shut off due to funding, prioritization, other hydrological or management reasons, then reestablished at a later time (several years, rather than weeks or months) at the same location. See also *discontinued record*. 28
- In a sample size of  $n$ , a known number of observations is missing at either end or at both ends (David, 1981; Cohen, 1991). 29

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

## 176 Guidelines for Determining Flood Flow Frequency – Bulletin 17C

1 A numerical measure or index of the lack of symmetry in a frequency distribution. Function of the third  
2 moment of magnitudes about their mean, a measure of asymmetry. Also called *coefficient of skew* or *skew*  
3 *coefficient*.

4 Computed values on both sides of an estimate of a parameter or quantile that show for a specified probability  
5 the range in which the true value of the parameter or quantile lies.

6 A simple, economical, reliable, and easily installed device for obtaining the elevation of the flood crest of  
7 streams (Sauer and Turnipseed, 2010). These gages are nonrecording and consist of a partial streamflow record.  
8 Flow intervals and perception thresholds are needed to describe each year of the flood record.

9 A measure of similarity, interdependence or relationship between two time series of observations in space  
10 at the same point or lagged points in time.

11 A systematic record where streamflow data collection and measurements have ceased. This typically  
12 occurs when a gage is shut off due to funding, prioritization, other hydrological or management reasons, and  
13 not reestablished. See also *broken record*.

14 Knowledge that the magnitude (discharge or stage) of a flood was larger than some level or threshold. For  
15 example, the flood exceeded the bridge deck.

16 The percentage of values that exceed a specified magnitude, 100 times exceedance probability.

17 Probability that a random event will exceed a specified magnitude in a given time period, usually one year  
18 unless otherwise indicated.

19 A generalized method of moments procedure to estimate the P-III distribution parameters using the  
20 entire data set, simultaneously employing regional skew information and a wide range of historical flood and  
21 threshold-exceedance information, while adjusting for any potentially influential low floods, missing values  
22 from an incomplete record, or zero flood years.

23 Those floods that are the largest magnitude at a gaging station or miscellaneous site and that substantially  
24 exceed the other flood observations (Costa and Jarrett, 2008).

25 The minimum stage or discharge level at a gaging station, below which observations are not recorded or  
26 published. Also called *base discharge*.

27 Streamflow data collected at streamflow-gaging stations. A gaging record can consist of systematic data and  
28 historical flood data.

29 A selected site on a stream equipped and operated to furnish basic data from which continuous, systematic  
30 records of stage and discharge may be obtained (Grover and Harrington, 1943; Rantz and Others, 1982a).

31 See *regional skew coefficient*.

32 Typically recent (hours to weeks) physical evidence of the (approximate) maximum flood stage (Jarrett and England, 2002).  
33 The physical evidence generally is of three types: (1) deposits along channel margins and in vegetation  
34 that consist of very light, floatable material such as pine needles, seeds, small twigs, grasses, and very  
35 fine sediments; (2) damage to vegetation such as bent or matted grasses, twigs, and branches, stripped  
36 leaves or bark; and (3) small erosional features such as scour lines. Benson and Dalrymple (1967) and  
37 Koenig and others (2016) discuss identification, preservation and rating of high-water marks. The HWM  
38 evidence is typically short-lived (weeks), but woody debris may last from several years to several decades  
39 in arid and semi-arid climates (Baker, 1987), and geomorphic evidence can be preserved for millennia. Physical  
40 evidence, such as marks on buildings and other structures, is also long-lived. This stage may represent a  
41 maximum discharge when a single-valued relationship exists between stage and discharge; Costa and Jarrett (2008) describe  
42 other hydraulic situations. See also *paleostage indicator*.

43 Broad category of data collected by humans prior to establishing systematic protocols; it generally consists  
44 of diaries, written accounts of settlements, folklore, and descriptions that may document periods where extreme  
45 weather and/or floods have occurred. It may also be used to infer times when there have been no large floods.  
46 These accounts were recorded in a manner that was preserved well enough that we know about it today.

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

Flood events which were directly observed by humans, generally in a non-systematic manner by non-hydrologists (Baker, 1987). These events usually occurred and were described in some qualitative and/or quantitative fashion prior to the systematic record. Information about the floods was recorded and preserved well enough so that we know about it today.

Records from the same populations. Floods may be from different populations because they occurred before the building of a dam and after the building of a dam, or before the watershed was urbanized and after it became urbanized, or because some are generated by summer storms and others by snowmelt, or because some were generated in El Nino years and some were in other years. It may be difficult in some cases to definitively say if the flood record is homogeneous.

A streamflow record in which some peak flows are missing because they were too low or high to record or the gage was out of operation for a short period because of flooding.

Floods whose magnitude are not known exactly, but are known to fall within a range or interval (Stedinger and others, 1983).

The probability of rejecting a hypothesis when it is in fact true. At a "10-percent" level of significance the probability is 1/10.

See outlier.

Sum of the squared differences between the true and estimated values of a quantity divided by the number of observations. It can also be defined as the bias squared plus the variance of the quantity (Stedinger and others, 1993).

A standard statistical computation for estimating the parameters of a distribution from the moments of the sample data.

A statistical test used to identify multiple potentially influential low flood observations in an annual maximum time series.

Knowledge that the magnitude (discharge or stage) of a flood was less than some level or threshold.

Outliers (extreme events) are observations that are exceedingly low or high compared to the distributional properties of the vast majority of the data. When plotted, along with a reasonably fitted CDF to the data, the outlier values plot far from the fitted line at the low or high ends of the distribution. A CDF, such as the LP-III, may not fit data sets with outliers, and the fitted curve usually fails to fit the bulk of the data as well as the outliers. Low outliers are outliers at the low end of the data set, near zero, at least in comparison with the rest of the data. On a log-probability plot, the low outliers impart a strong downward curvature and a downward-drooping lower tail to the frequency curve. In comparison with the lower tail, the upper tail of the low-outlier-affected curve may appear relatively flat.

Physical evidence of past floods and their ages as observed from the geologic record or from botanical evidence. Paleoflood data typically consists of observations on individual past floods such as those derived from slackwater deposits, boulder bars, silt lines, or botanical information, that are collected as part of a paleoflood hydrology study (Benito and O'Connor, 2013). It can also consist of periods of landscape stability that can be used to place limits on flood magnitude over time, such as paleohydrologic bounds (Levish, 2002). Paleoflood data are distinguished from historical flood data, as a separate line of evidence, by the use of applied field geology techniques to examine and describe the geomorphic and stratigraphic context of extreme floods. In some cases, there is overlap between historical and paleoflood data, as historical and cultural artifacts such as barbed wire, beer cans (House and Baker, 2001) or pottery may be observed and used in dating and estimation of floods. The study of past or ancient floods which occurred prior to the time of human observation or direct measurement by modern hydrologic procedures (Baker, 1987). Paleoflood hydrology has also been defined as "the study of the movements of water and sediment in channels before the time of continuous hydrologic records or direct measurements" (Costa, 1986).

A time interval during which a given discharge has not been exceeded (Levish, 2002). The term is sometimes shortened to *bound*. The paleohydrologic bound represents stages and discharges that have not been

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

## 178 Guidelines for Determining Flood Flow Frequency – Bulletin 17C

1 exceeded since the geomorphic surface stabilized. Bounds are appropriate for paleohydrologic information  
2 and are not dependent on human observation of a particular event, but on the physical setting (hydraulic and  
3 geomorphic). (alt: *paleoflood bound*).–

4 An erosional or depositional feature that recorded the near peak stage of an individual flood (Jarrett and England, 2002) prior  
5 to human observation. Indirect evidence of the stage of past floods includes botanical evidence and sedimentological  
6 deposits (Jarrett, 1991). Large floods, especially in high gradient channels, can transport and deposit coarse  
7 material (gravel, boulders, and woody debris, etc.) that may be interpreted as HWMs. The primary differences  
8 between PSIs and HWMs are: (1) HWMs represent events which occurred “more recent” in time due to their  
9 relatively short preservation length as compared to PSIs; and (2) some PSIs may not represent the exact peak  
10 stage.–

11 A list of all flows (such as flood peaks) that exceed a chosen base stage or discharge, regardless of  
12 the number of peaks occurring in a year. Also called basic-stage flood series, or floods above a base  
13 (Langbein and Iseri, 1960).–

14 A probability multiplied by 100.–

15 The stage or flow above which it is estimated a source would provide information on the flood peak  
16 in any given year. Perception thresholds ( $T_{Y,lower}$ ;  $T_{Y,upper}$ ) reflect the range of flows that would have been  
17 measured/recorded had they occurred. If an event magnitude had occurred in a specific year, there is  
18 information indicate it would have been “recorded” in a manner that we could perceive it today. Perception  
19 thresholds describe the range of measurable potential discharges and are independent of the actual peak  
20 discharges that have occurred. They are used to provide a rank and record length for each reported flood peak  
21 (Gerard and Karpuk, 1979). Perception thresholds are used for historical data, when the information provided  
22 is based on human observation. They are also used to describe a paleoflood period and paleoflood data. In  
23 addition, perception thresholds are used to properly accommodate unrecorded floods below a “gage base”. A  
24 perception threshold is allocated to each information source for each year  $Y$  of the flood record. Perception  
25 thresholds may involve a significant amount of judgment on the part of the scientist and/or historian regarding,  
26 for any given year, what is the smallest event that would have been recorded (in a physical or textural manner)  
27 such that we would actually know about it today (alt: *threshold*).–

28 The entire (usually infinite) number of data from which a sample is taken or collected. The total number of  
29 past, present, and future floods at a location on a river is the population of floods for that location even if the  
30 floods are not measured or recorded.–

31 In an annual maximum flood series, small-magnitude flows (including zeros) that do not represent the  
32 physical processes that cause the largest flood observations. These “PILFs” can exert high leverage and influence  
33 on the flood frequency distribution.–

34 Estimate of the flood magnitude  $Q$  for exceedance probability  $p$  from a fitted distribution.–

35 A procedure to improve the accuracy of the moments (mean and variance) of a short-record flood series by  
36 using information from longer records at nearby locations with high cross-correlation (Matalas and Jacobs, 1964; Stedinger and others, 1993).–

37  
38 The creation of a longer flood-flow record (individual floods) at a site with a short record, by using flood  
39 observations at a long-record site with high cross-correlation. The technique can also be used to fill in missing  
40 observations (Hirsch and others, 1993; Stedinger and others, 1993).–

41 A skew coefficient derived by a procedure which integrates values obtained at many locations.–

42 In flood frequency, a procedure that is reasonably efficient when the assumed characteristics of the flood  
43 distribution are true, while not doing poorly when those assumptions are violated (Kuczera, 1982; Cohn and others, 2013).–

44  
45 An element, part, or fragment of a “population.” Every hydrologic record is a sample of a much longer  
46 record.–

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017

~~See autocorrelation.~~

~~See coefficient of skewness.~~

~~A measure of the dispersion or precision, of a series of statistical values such as precipitation or streamflow. It is the square root of the sum of squares of the deviations from the arithmetic mean divided by the number of values or events in the series. It is standard practice to divide by the number of values minus one in order to get an unbiased estimate of the variance from the sample data.~~

~~An estimate of the standard deviation of a statistic. Often calculated from a single set of observations. Calculated like the standard deviation but differing from it in meaning.~~

~~Data that are collected at regular, prescribed intervals under a defined protocol. In the context of streamflow, systematic data consist of discharge and stage data collected at regular, prescribed intervals, typically at gaging stations. (syn. systematic record).~~

~~See perception threshold.~~

~~A measure of the amount of spread or dispersion of a set of values around their mean, obtained by calculating the mean value of the squares of the deviations from the mean, and hence equal to the square of the standard deviation.~~

~~A value obtained by multiplying each of a series of values by its assigned weight and dividing the sum of those products by the sum of the weights.~~

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

DRAFT: August 31, 2017



DRAFT: August 31, 2017

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

Prepared by John F. England, Jr., William H. Asquith and the USGS Reston Publishing Service Center  
Edited by Marilyn A. Billone, USGS  
Illustrations by Authors and USGS  
Layout and typesetting by John F. England using  $\text{\LaTeX}$   
Final portable document format preparation by William H. Asquith, USGS

For information regarding this report , contact:

U.S. Geological Survey  
415 National Center  
12201 Sunrise Valley Drive  
Reston, Virginia 20192  
(703) 648-5301  
<https://water.usgs.gov/osw/>

DRAFT: August 31, 2017

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.

1

2

DRAFT: August 31, 2017

---PROVISIONAL---

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE  
PURPOSE OF OBTAINING PEER REVIEW UNDER THE USGS  
PEER REVIEW PLAN.

IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE U.S.  
GEOLOGICAL SURVEY (USGS).

IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED  
TO REPRESENT ANY OFFICIAL USGS FINDINGS OR POLICY.